GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2005, 14:09:02 ; Search time 137 Seconds
(without alignments)
713.263 Million cell updates/sec

Title: US-10-809-655-9
Perfect score: 1528
Sequence: 1 MSPRGTGCSAGLLMTVGWLL.....LDQLPTEMPGEDDALSEWNE 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1421835 seqs, 332370683 residues

Total number of hits satisfying chosen parameters: 1421835

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1528	100.0	294	14	US-10-181-642-9	Sequence 9, Appli
2	1528	100.0	294	16	US-10-809-654-9	Sequence 9, Appli
3	1528	100.0	294	16	US-10-809-655-9	Sequence 9, Appli
4	762.5	49.9	163	14	US-10-411-224-105	Sequence 105, App
5	762.5	49.9	163	15	US-10-047-021-105	Sequence 105, App
6	593	38.8	124	15	US-10-047-021-119	Sequence 119, App
7	593	38.8	125	14	US-10-411-224-119	Sequence 119, App
8	110	7.2	1315	15	US-10-369-493-3346	Sequence 3346, Ap
9	103	6.7	255	11	US-09-833-245-1568	Sequence 1568, Ap
10	100	6.5	604	16	US-10-437-963-174055	Sequence 174055,
11	99	6.5	255	11	US-09-833-245-1567	Sequence 1567, Ap
12	99	6.5	255	13	US-10-052-586-538	Sequence 538, App
13	99	6.5	255	14	US-10-174-590-538	Sequence 538, App
14	99	6.5	255	14	US-10-176-758-538	Sequence 538, App
15	99	6.5	255	14	US-10-175-737-538	Sequence 538, App
16	99	6.5	255	14	US-10-174-581-538	Sequence 538, App
17	99	6.5	255	14	US-10-176-483-538	Sequence 538, App
18	99	6.5	255	14	US-10-176-749-538	Sequence 538, App
19	99	6.5	255	14	US-10-176-914-538	Sequence 538, App
20	99	6.5	255	14	US-10-176-915-538	Sequence 538, App
21	99	6.5	255	14	US-10-173-706-538	Sequence 538, App
22	99	6.5	255	14	US-10-175-738-538	Sequence 538, App
23	99	6.5	255	14	US-10-175-752-538	Sequence 538, App
24	99	6.5	255	14	US-10-176-482-538	Sequence 538, App
25	99	6.5	255	14	US-10-176-757-538	Sequence 538, App
26	99	6.5	255	14	US-10-176-913-538	Sequence 538, App
27	99	6.5	255	14	US-10-180-552-538	Sequence 538, App
28	99	6.5	255	14	US-10-180-557-538	Sequence 538, App
29	99	6.5	255	14	US-10-173-700-538	Sequence 538, App
30	99	6.5	255	14	US-10-174-572-538	Sequence 538, App
31	99	6.5	255	14	US-10-174-579-538	Sequence 538, App
32	99	6.5	255	14	US-10-174-582-538	Sequence 538, App
33	99	6.5	255	14	US-10-174-588-538	Sequence 538, App
34	99	6.5	255	14	US-10-175-739-538	Sequence 538, App
35	99	6.5	255	14	US-10-175-740-538	Sequence 538, App
36	99	6.5	255	14	US-10-175-743-538	Sequence 538, App
37	99	6.5	255	14	US-10-176-488-538	Sequence 538, App
38	99	6.5	255	14	US-10-176-492-538	Sequence 538, App
39	99	6.5	255	14	US-10-176-747-538	Sequence 538, App
40	99	6.5	255	14	US-10-176-750-538	Sequence 538, App
41	99	6.5	255	14	US-10-176-985-538	Sequence 538, App
42	99	6.5	255	14	US-10-176-987-538	Sequence 538, App
43	99	6.5	255	14	US-10-176-992-538	Sequence 538, App
44	99	6.5	255	14	US-10-176-993-538	Sequence 538, App
45	99	6.5	255	14	US-10-184-658-538	Sequence 538, App

ALIGNMENTS

RESULT 1
US-10-181-642-9
; Sequence 9, Application US/10181642
; Publication No. US20030087375A1
; GENERAL INFORMATION:
; APPLICANT: Herr, John C.
; APPLICANT: Shetty, Jagathapala
; APPLICANT: Wolkowicz, Michael
; APPLICANT: Jayes, Friederike
; APPLICANT: Hao, Zhonglin
; TITLE OF INVENTION: Sperm Specific Proteins
; FILE REFERENCE: 00497-02
; CURRENT APPLICATION NUMBER: US/10/181,642
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/176,885
; PRIOR FILING DATE: 2000-01-19
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-181-642-9

Parent

Query Match 100.0%; Score 1528; DB 14; Length 294;
Best Local Similarity 100.0%; Pred. No. 5e-150;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
```

```
Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180

Qy        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240

Qy        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294


RESULT 2
US-10-809-654-9
; Sequence 9, Application US/10809654
; Publication No. US20040161824A1
; GENERAL INFORMATION:
;  APPLICANT: Herr, John C.
;  APPLICANT:  Shetty, Jagathapala
;  APPLICANT:  Wolkowicz, Michael
;  APPLICANT:  Jayes, Friederike
;  APPLICANT:  Hao, Zhonglin
;  TITLE OF INVENTION: Sperm Specific Proteins
;  FILE REFERENCE: 00497-02
;  CURRENT APPLICATION NUMBER: US/10/809,654
;  CURRENT FILING DATE:  2004-03-25
;  PRIOR APPLICATION NUMBER: 60/176,885
;  PRIOR FILING DATE: 2000-01-19
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 294
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-809-654-9

  Query Match           100.0%;  Score 1528;  DB 16;  Length 294;
  Best Local Similarity  100.0%;  Pred. No. 5e-150;
  Matches  294;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180

Qy        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240

Qy        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294


RESULT 3
US-10-809-655-9
; Sequence 9, Application US/10809655
; Publication No. US20040161825A1
; GENERAL INFORMATION:
;  APPLICANT: Herr, John C.
;  APPLICANT:  Shetty, Jagathapala
;  APPLICANT:  Wolkowicz, Michael
;  APPLICANT:  Jayes, Friederike
;  APPLICANT:  Hao, Zhonglin
;  TITLE OF INVENTION: Sperm Specific Proteins
;  FILE REFERENCE: 00497-02
;  CURRENT APPLICATION NUMBER: US/10/809,655
;  CURRENT FILING DATE:  2004-03-25
;  PRIOR APPLICATION NUMBER: 60/176,885
;  PRIOR FILING DATE: 2000-01-19
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 294
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-809-655-9

  Query Match           100.0%;  Score 1528;  DB 16;  Length 294;
  Best Local Similarity  100.0%;  Pred. No. 5e-150;
  Matches  294;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180

Qy        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240

Qy        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294


RESULT 4
US-10-411-224-105
; Sequence 105, Application US/10411224
; Publication No. US20030166906A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 50 Human Secreted Proteins
;  FILE REFERENCE: PZ016P1
;  CURRENT APPLICATION NUMBER: US/10/411,224
;  CURRENT FILING DATE:  2003-04-11
;  PRIOR APPLICATION NUMBER: US/09/722,329
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: 09/262,109
;  PRIOR FILING DATE: 1999-03-04
;  PRIOR APPLICATION NUMBER: 60/057,626
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: 60/057,663
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: 60/057,669
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: 60/058,667
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,974
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,973
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,666
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/090,112
;  PRIOR FILING DATE: 1998-06-22
;  NUMBER OF SEQ ID NOS: 206
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 105
;   LENGTH: 163
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (106)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (113)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-411-224-105

  Query Match             49.9%;  Score 762.5;  DB 14;  Length 163;
  Best Local Similarity   94.2%;  Pred. No. 9e-71;
  Matches  147;  Conservative    2;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              |||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGEXKCVVRVXECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQ 156
              |||||||||||||||||||||||    |: :||||
Db        121 CGWGKPISESLESVRLACIHTSPLIVSIYV-ELLRQ 155


RESULT 5
US-10-047-021-105
; Sequence 105, Application US/10047021
; Publication No. US20040002591A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 50 Human Secreted Proteins
;  FILE REFERENCE: PZ016P2
;  CURRENT APPLICATION NUMBER: US/10/047,021
;  CURRENT FILING DATE:  2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/262,066
;  PRIOR FILING DATE: 2001-01-18
;  PRIOR APPLICATION NUMBER: US 09/722,329
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: US 09/262,109
;  PRIOR FILING DATE: 1999-03-04
;  PRIOR APPLICATION NUMBER: PCT/US98/18360
;  PRIOR FILING DATE: 1998-09-03
;  PRIOR APPLICATION NUMBER: US 60/057,626
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: US 60/057,663
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: US 60/057,669
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: US 60/058,667
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/058,974
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/058,973
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/058,666
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/090,112
;  PRIOR FILING DATE: 1998-06-22
;  NUMBER OF SEQ ID NOS: 206
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 105
;   LENGTH: 163
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (106)
;   OTHER INFORMATION: Xaa equals any amino acid
;   NAME/KEY: SITE
;   LOCATION: (113)
;   OTHER INFORMATION: Xaa equals any amino acid
US-10-047-021-105

  Query Match             49.9%;  Score 762.5;  DB 15;  Length 163;
  Best Local Similarity   94.2%;  Pred. No. 9e-71;
  Matches  147;  Conservative    2;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              |||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGEXKCVVRVXECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQ 156
              |||||||||||||||||||||||    |: :||||
Db        121 CGWGKPISESLESVRLACIHTSPLIVSIYV-ELLRQ 155


RESULT 6
US-10-047-021-119
; Sequence 119, Application US/10047021
; Publication No. US20040002591A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 50 Human Secreted Proteins
;  FILE REFERENCE: PZ016P2
;  CURRENT APPLICATION NUMBER: US/10/047,021
;  CURRENT FILING DATE:  2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/262,066
;  PRIOR FILING DATE: 2001-01-18
;  PRIOR APPLICATION NUMBER: US 09/722,329
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: US 09/262,109
;  PRIOR FILING DATE: 1999-03-04
;  PRIOR APPLICATION NUMBER: PCT/US98/18360
;  PRIOR FILING DATE: 1998-09-03
;  PRIOR APPLICATION NUMBER: US 60/057,626
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: US 60/057,663
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: US 60/057,669
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: US 60/058,667
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/058,974
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/058,973
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/058,666
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: US 60/090,112
;  PRIOR FILING DATE: 1998-06-22
;  NUMBER OF SEQ ID NOS: 206
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
;   LENGTH: 124
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (75)
;   OTHER INFORMATION: Xaa equals any amino acid
US-10-047-021-119

  Query Match             38.8%;  Score 593;  DB 15;  Length 124;
  Best Local Similarity   98.3%;  Pred. No. 2.6e-53;
  Matches  113;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
              |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPSETEDVSNRNXVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115


RESULT 7
US-10-411-224-119
; Sequence 119, Application US/10411224
; Publication No. US20030166906A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 50 Human Secreted Proteins
;  FILE REFERENCE: PZ016P1
;  CURRENT APPLICATION NUMBER: US/10/411,224
;  CURRENT FILING DATE:  2003-04-11
;  PRIOR APPLICATION NUMBER: US/09/722,329
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: 09/262,109
;  PRIOR FILING DATE: 1999-03-04
;  PRIOR APPLICATION NUMBER: 60/057,626
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: 60/057,663
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: 60/057,669
;  PRIOR FILING DATE: 1997-09-05
;  PRIOR APPLICATION NUMBER: 60/058,667
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,974
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,973
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,666
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/090,112
;  PRIOR FILING DATE: 1998-06-22
;  NUMBER OF SEQ ID NOS: 206
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
;   LENGTH: 125
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (75)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (125)
;   OTHER INFORMATION: Xaa equals stop translation
US-10-411-224-119

  Query Match             38.8%;  Score 593;  DB 14;  Length 125;
  Best Local Similarity   98.3%;  Pred. No. 2.6e-53;
  Matches  113;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
              |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPSETEDVSNRNXVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115


RESULT 8
US-10-369-493-3346
; Sequence 3346, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3346
;   LENGTH: 1315
;   TYPE: PRT
;   ORGANISM: Neurospora crassa
US-10-369-493-3346

  Query Match              7.2%;  Score 110;  DB 15;  Length 1315;
  Best Local Similarity   25.2%;  Pred. No. 0.14;
  Matches   61;  Conservative   38;  Mismatches   93;  Indels   50;  Gaps   14;

Qy         41 LAHEGEGEEETENNDSETAENYA---PPETEDV---SNRNVVKEVEFGMCTVTCGIGVRE 94
              ||||    :||| :    |         : :||   |::|    |
Db        378 LAHESAANKETEEQEVPAKEEMLLLDHVDDDDVGEGSSKNAKTSKESN----------- 425

Qy         95 VILTNGCPGGES-KCVVRVEECRGPTDCGW----GKPISESL---ESVRLACIHTSP--- 143
              ||  |  | :  :| ::  | ||  |       |:|::|    | |||       |
Db        426 ---NNGAEGSSSGETFIRTDQLDGETD--WKLRLASPLSQNLSTEELVRLRVTAGKPDKR 480

Qy        144 LNRFKYMWKLL--RQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA----TIKF 197
              :| |    :||  |||  |     |     :|:  : ||: : :|   | ::|    |:
Db        481 VNEFVGTLELLPSRQDVMSGAAYNPREGDDVK--AAPLSID-NTAWANTVIASNATTLAV 537

Qy        198 TVYTSSELQMRRSSLPATDAALIF---VLTIGVIICVFIIFLLIFIIINWAAVKAFWGAK 254
              ||| : :   |: |:     :    : ::  |:|    || :|: |   |:: |  ||
Db        538 IVYTGPQTRSALSTSPSRSKTGLLEYEINSLTKILC----FLTLFLSIVLVALEGFSTAK 593

Qy        255 AS 256
              ::
Db        594 GN 595


RESULT 9
US-09-833-245-1568
; Sequence 1568, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Albumin Fusion Proteins
;  FILE REFERENCE: PF546PCT
;  CURRENT APPLICATION NUMBER: US/09/833,245
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/229, 358
;  PRIOR FILING DATE: 2000-04-12
;  PRIOR APPLICATION NUMBER: 60/256, 931
;  PRIOR FILING DATE: 2000-12-21
;  PRIOR APPLICATION NUMBER: 60/199, 384
;  PRIOR FILING DATE: 2000-04-25
;  NUMBER OF SEQ ID NOS: 2267
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1568
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-245-1568

  Query Match              6.7%;  Score 103;  DB 11;  Length 255;
  Best Local Similarity   24.8%;  Pred. No. 0.069;
  Matches   41;  Conservative   21;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              | :  | ||    :   |:  :    ||||||:| : | :   |  |   ||  |  ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTRRLECL 82
```

```
Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 10
US-10-437-963-174055
; Sequence 174055, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 174055
;   LENGTH: 604
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(604)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_72033C.1.pep
US-10-437-963-174055

  Query Match             6.5%;  Score 100;  DB 16;  Length 604;
  Best Local Similarity   24.6%;  Pred. No. 0.5;
  Matches   51;  Conservative   27;  Mismatches   63;  Indels   66;  Gaps   12;

Qy         16 VGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNV 75
Db        304 IGMFRKAALEDVK--RFKEAVMEIG---EEEGEEDDEKKDK----------EDDGLRGL 347

Qy         76 VKEVEFGMCTVTCGIGVREVILTN-----GCP--GGES------KCVVRVEECRGPTDCG 122
Db        348 SGSIFQGPLVYSC-LVVRSILRTSCLGILGFPFVGASSVVNSNLPCSSRISSCSDFTSGY 406

Qy        123 WGKPI-------SESLESVRLACIHTSPL-NRFKYMWKLLRQDQQSIILV---------- 164
Db        407 SWRPIEAAKLCRTRAVRSLQITCTATKPAKSPAEEEWKIKRQ-----VLVEKRVRSVDVK 461

Qy        165 --------NDSAILEVR-----KESHP 178
Db        462 EALRLQKENNPAILDVRPVADFKEAHP 488


RESULT 11
US-09-833-245-1567
; Sequence 1567, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Albumin Fusion Proteins
;  FILE REFERENCE: PF546PCT
;  CURRENT APPLICATION NUMBER: US/09/833,245
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/229, 358
;  PRIOR FILING DATE: 2000-04-12
;  PRIOR APPLICATION NUMBER: 60/256, 931
;  PRIOR FILING DATE: 2000-12-21
;  PRIOR APPLICATION NUMBER: 60/199, 384
;  PRIOR FILING DATE: 2000-04-25
;  NUMBER OF SEQ ID NOS: 2267
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1567
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-245-1567

  Query Match             6.5%;  Score 99;  DB 11;  Length 255;
  Best Local Similarity   24.2%;  Pred. No. 0.18;
  Matches   40;  Conservative   22;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTQRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 12
US-10-052-586-538
; Sequence 538, Application US/10052586
; Publication No. US20020127584A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C1
;  CURRENT APPLICATION NUMBER: US/10/052,586
;  CURRENT FILING DATE:  2002-01-15
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059266
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/063120
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063121
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063486
;  PRIOR FILING DATE: 1997-10-21
;  PRIOR APPLICATION NUMBER: 60/063540
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063541
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063544
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063564
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063734
;  PRIOR FILING DATE: 1997-10-29
```

; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088722
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088740
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12

```
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089512
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089538
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089598
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089653
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089908

  Query Match           6.5%;  Score 99;  DB 13;  Length 255;
  Best Local Similarity  24.2%;  Pred. No. 0.18;
  Matches   40;  Conservative   22;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              | :  | ||    :   |:  :    ||||||:| : | :   |  |   ||  :  ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTQRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                 ||         ||          |:|: :  |      |:: |:| |     :|  :
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
                |      :  || : |:          ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 13
US-10-174-590-538
; Sequence 538, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C42
;  CURRENT APPLICATION NUMBER: US/10/174,590
;  CURRENT FILING DATE:  2002-06-18
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 538
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-174-590-538

  Query Match           6.5%;  Score 99;  DB 14;  Length 255;
  Best Local Similarity  24.2%;  Pred. No. 0.18;
  Matches   40;  Conservative   22;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              | :  | ||    :   |:  :    ||||||:| : | :   |  |   ||  :  ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTQRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                 ||         ||          |:|: :  |      |:: |:| |     :|  :
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
                |      :  || : |:          ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 14
US-10-176-758-538
; Sequence 538, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C104
;  CURRENT APPLICATION NUMBER: US/10/176,758
;  CURRENT FILING DATE:  2002-06-21
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 538
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-758-538

  Query Match           6.5%;  Score 99;  DB 14;  Length 255;
  Best Local Similarity  24.2%;  Pred. No. 0.18;
  Matches   40;  Conservative   22;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              | :  | ||    :   |:  :    ||||||:| : | :   |  |   ||  :  ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTQRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                 ||         ||          |:|: :  |      |:: |:| |     :|  :
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
                |      :  || : |:          ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 15
US-10-175-737-538
; Sequence 538, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C50
;  CURRENT APPLICATION NUMBER: US/10/175,737
;  CURRENT FILING DATE:  2002-06-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
```

; SEQ ID NO 538
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-538

Query Match 6.5%; Score 99; DB 14; Length 255;
Best Local Similarity 24.2%; Pred. No. 0.18;
Matches 40; Conservative 22; Mismatches 65; Indels 38; Gaps 8;

```
Qy          58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
               | :  | ||    :   |:  :    |||||||: | : |  :   |  |   ||  :  ||
Db          25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTQRLECL 82

Qy         117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                  ||         ||          |:|: :  |      |:: |:| |     :|  :
Db          83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy         166 DSAILEVRKESHPLAFE----------CDT--LDNNEIVATIKF 197
               |      :  || : |:          ||   : |  :|  : |
Db         132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176
```

Search completed: April 15, 2005, 14:18:51
Job time : 138 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 15, 2005, 14:04:21 ; Search time 173 Seconds
(without alignments)
657.269 Million cell updates/sec

Title: US-10-809-655-9
Perfect score: 1528
Sequence: 1 MSPRGTGCSAGLLMTVGWLL.....LDQLPTEMPGEDDALSEWNE 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1528	100.0	294	4	AAB85346	Aab85346 Sperm spe
2	762.5	49.9	163	5	ABP62052	Abp62052 Human sec
3	762.5	49.9	163	6	ADA57229	Ada57229 Human sec
4	762.5	49.9	163	6	ADA41111	Ada41111 Human sec
5	762.5	49.9	163	7	ADD37978	Add37978 Human sec
6	593	38.8	124	5	ABP62066	Abp62066 Human sec
7	593	38.8	124	6	ADA57562	Ada57562 Human sec
8	593	38.8	124	6	ADA41456	Ada41456 Human sec
9	593	38.8	124	7	ADD38087	Add38087 Human sec
10	593	38.8	125	2	AAY12953	Aay12953 Amino aci
11	110	7.2	1315	8	ADN20693	Adn20693 Bacterial
12	103	6.7	255	2	AAY31833	Aay31833 Human foe
13	103	6.7	255	4	ABG02357	Abg02357 Novel hum
14	103	6.7	255	5	AAU96170	Aau96170 Human sec
15	103	6.7	255	5	ABG64819	Abg64819 Human alb
16	103	6.7	255	8	ADL78086	Adl78086 Albumin f
17	102	6.7	1083	4	ABB61710	Abb61710 Drosophil
18	99	6.5	255	4	AAB73684	Aab73684 Human oxi
19	99	6.5	255	4	AAU29292	Aau29292 Human PRO
20	99	6.5	255	5	AAU96195	Aau96195 Human sec
21	99	6.5	255	5	ABG64818	Abg64818 Human alb
22	99	6.5	255	6	ABU58668	Abu58668 Human PRO
23	99	6.5	255	6	ABU88216	Abu88216 Novel hum
24	99	6.5	255	6	ABU84531	Abu84531 Human sec
25	99	6.5	255	6	ABR66405	Abr66405 Human sec
26	99	6.5	255	6	ABR65795	Abr65795 Human sec
27	99	6.5	255	6	ABU99735	Abu99735 Human sec
28	99	6.5	255	6	ABU82974	Abu82974 Human PRO
29	99	6.5	255	6	ABU90095	Abu90095 Novel hum
30	99	6.5	255	6	ABR68344	Abr68344 Human sec
31	99	6.5	255	6	ABU96397	Abu96397 Novel hum
32	99	6.5	255	6	ABU92828	Abu92828 Human sec
33	99	6.5	255	6	ABO08905	Abo08905 Human sec
34	99	6.5	255	6	ABO02957	Abo02957 Human sec
35	99	6.5	255	6	ABR75111	Abr75111 Human sec
36	99	6.5	255	6	ABR94873	Abr94873 Human sec
37	99	6.5	255	6	ABU85846	Abu85846 Human PRO
38	99	6.5	255	6	ABU99006	Abu99006 Novel hum
39	99	6.5	255	6	ABU98221	Abu98221 Novel hum
40	99	6.5	255	6	ABU91927	Abu91927 Novel hum
41	99	6.5	255	6	ABU89620	Abu89620 Human PRO
42	99	6.5	255	6	ABU86461	Abu86461 Human sec
43	99	6.5	255	6	ABU67674	Abu67674 Human sec
44	99	6.5	255	6	ABU80702	Abu80702 Human PRO
45	99	6.5	255	6	ABR99620	Abr99620 Human sec

ALIGNMENTS

RESULT 1
AAB85346
ID AAB85346 standard; protein; 294 AA.
XX
AC AAB85346;
XX
DT 17-SEP-2001 (first entry)
XX
DE Sperm specific surface protein SAMP32.
XX
KW Sperm specific surface protein; C7/8; SAMP32; C58; contraceptive;
KW vaccine; antifertility; spermicide.
XX
OS Homo sapiens.
XX
PN WO200153352-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US001717.
XX
PR 19-JAN-2000; 2000US-0176885P.
XX
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI Hao Z, Herr JC, Jayes FL, Shetty J, Wolkowicz MJ;
XX
DR WPI; 2001-451902/48.
DR N-PSDB; AAH22947.
XX
PT New human sperm surface proteins C7/8, SAMP32 and C58 for development of
PT contraceptive vaccines.
XX
PS Claim 10; Page 56-67; 63pp; English.
XX
CC The invention relates to novel human sperm specific surface proteins,
CC C7/8, SAMP32 and C58. The proteins, nucleic acids encoding the sperm
CC specific surface proteins and antibodies specific for the proteins are
CC useful for making contraceptive compositions including contraceptive
CC vaccines. The vaccines produced by the invention are hoped to be more
CC effective than previous contraceptive vaccines which only caused 75%
CC inhibition of fertility. The present sequence represents the human sperm
CC specific surface protein SAMP32
XX
SQ Sequence 294 AA;

Query Match 100.0%; Score 1528; DB 4; Length 294;
Best Local Similarity 100.0%; Pred. No. 3.2e-157;

```
  Matches  294;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180

Qy        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240

Qy        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294


RESULT 2
ABP62052
ID   ABP62052 standard; protein; 163 AA.
XX
AC   ABP62052;
XX
DT   12-NOV-2002  (first entry)
XX
DE   Human secreted protein SEQ ID NO 105.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200257420-A2.
XX
PD   25-JUL-2002.
XX
PF   17-JAN-2002; 2002WO-US001109.
XX
PR   18-JAN-2001; 2001US-0262066P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Moore PA,  Ruben SM,  Lafleur DW,  Shi Y,  Rosen CA,  Olsen H;
PI   Ebner R,  Brewer LA;
XX
DR   WPI; 2002-599716/64.
DR   N-PSDB; ABQ92592.
XX
PT   New polynucleotides and polypeptides useful for diagnosing, prognosing,
PT   treating or preventing e.g. neurodegenerative, central nervous system,
PT   autoimmune, respiratory, reproductive, or inflammatory diseases or
PT   disorders.
XX
PS   Claim 11; Page 749-750; 785pp; English.
XX
CC   The invention relates to novel genes (ABQ92553-ABQ92607) and proteins
CC   (ABP62013-ABP62153) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections
XX
SQ   Sequence 163 AA;

  Query Match           49.9%;  Score 762.5;  DB 5;  Length 163;
  Best Local Similarity  94.2%;  Pred. No. 3.2e-74;
  Matches  147;  Conservative    2;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGEXKCVVRVXECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQ 156
              ||||||||||||||||||||||    |: :||||
Db        121 CGWGKPISESLESVRLACIHTSPLIVSIYV-ELLRQ 155


RESULT 3
ADA57229
ID   ADA57229 standard; protein; 163 AA.
XX
AC   ADA57229;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein #512.
XX
KW   immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
KW   cytostatic; cerebroprotective; neuroprotective; nootropic;
KW   cardiovascular; antiarteriosclerotic; gene therapy;
KW   human secreted protein; immune disorder; inflammation;
KW   respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
KW   inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
KW   multiple sclerosis; ischaemic brain injury; Parkinson's disease;
KW   Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
KW   triple helix formation; antisense gene therapy; forensic biology.
XX
OS   Homo sapiens.
XX
PN   WO2002102994-A2.
XX
PD   27-DEC-2002.
XX
PF   19-MAR-2002; 2002WO-US008278.
XX
PR   21-MAR-2001; 2001US-0277340P.
PR   19-JUL-2001; 2001US-0306171P.
PR   13-NOV-2001; 2001US-0331287P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-167512/16.
DR   N-PSDB; ADA56333.
XX
PT   New human secreted polypeptides and polynucleotides, useful for
PT   diagnosing, treating or preventing e.g. immune disorders, inflammatory
PT   conditions, respiratory disorders, cancers, CNS disorders, or
PT   neurodegenerative disorders.
XX
PS   Claim 13; SEQ ID NO 1419; 1754pp; English.
```

```
XX
CC   The invention relates to 592 new human secreted polypeptides useful for
CC   diagnosing, treating or preventing e.g. immune disorders, inflammatory
CC   conditions, respiratory disorders, cancers, CNS disorders, or
CC   neurodegenerative disorders, or polypeptides comprising an amino acid
CC   sequence at least 95% identical to the new sequences. The polypeptides,
CC   antibodies or antibody fragments that bind to the polypeptides, nucleic
CC   acids encoding the polypeptides, agonists or antagonists that binds to
CC   the polypeptide, are useful in preparing diagnostic or pharmaceutical
CC   compositions for diagnosing, treating or preventing an e.g. immune
CC   disorders, inflammatory conditions (e.g. inflammatory bowel disease,
CC   nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
CC   allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
CC   (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
CC   disorders (e.g. Parkinson's disease or Alzheimer's disease), and
CC   cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
CC   polynucleotides are useful for chromosome identification, chromosome
CC   mapping, for controlling gene expression through triple helix formation
CC   or antisense DNA or RNA, in gene therapy, for identifying individuals
CC   from minute biological samples, in forensic biology, and as hybridization
CC   probes. The polypeptides are useful for as molecular weight markers on
CC   sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
CC   gels, to raise antibodies, for testing biological activities, and for
CC   treating or preventing neural disorders, immune system disorders,
CC   muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC   renal, proliferative and/or cancerous diseases. This sequence corresponds
CC   to one of the polypeptide of the invention. Note: The sequence data for
CC   this patent did form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 163 AA;

  Query Match             49.9%;  Score 762.5;  DB 6;  Length 163;
  Best Local Similarity   94.2%;  Pred. No. 3.2e-74;
  Matches  147;  Conservative    2;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              |||||||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGEXKCVVRVXECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQ 156
              |||||||||||||||||||||||    |: :||||
Db        121 CGWGKPISESLESVRLACIHTSPLIVSIYV-ELLRQ 155


RESULT 4
ADA41111
ID   ADA41111 standard; protein; 163 AA.
XX
AC   ADA41111;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein.
XX
KW   Human; secreted protein; cancer; hyperproliferative disorder;
KW   rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW   anaemia; allergic reaction; asthma; cardiovascular disorder;
KW   wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective;
KW   antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW   vulnerary; cardiant; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO2002102993-A2.
XX
PD   27-DEC-2002.
XX
PF   19-MAR-2002; 2002WO-US008123.
XX
PR   21-MAR-2001; 2001US-0277340P.
PR   19-JUL-2001; 2001US-0306171P.
PR   13-NOV-2001; 2001US-0331287P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-175238/17.
XX
PT   New human secreted proteins and nucleic acid molecules, useful for
PT   preparing a diagnostic or pharmaceutical composition for diagnosing,
PT   preventing or treating cancer or other hyperproliferative disorder,
PT   asthma, allergies or AIDS.
XX
PS   Claim 1; SEQ ID NO 1493; 3205pp; English.
XX
CC   The invention relates to novel genes ADA39629-ADA40565 and proteins
CC   ADA40566-ADA41501 for human secreted proteins, useful for preventing,
CC   treating or ameliorating medical conditions e.g. by protein or gene
CC   therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC   fragments, and agonists or antagonists that bind to the polypeptide are
CC   useful for preparing a diagnostic or pharmaceutical composition for
CC   diagnosing or treating cancer or other hyperproliferative disorder. The
CC   polypeptides and nucleic acid molecules are also useful for detecting,
CC   preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC   or other hyperproliferative disorders including neoplasms, autoimmune
CC   disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC   erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC   anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC   thrombocytopenia), allergic reactions including asthma or eczema,
CC   inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC   bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC   Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC   (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC   fungal or viral infections including HIV/AIDS), or wound healing and
CC   disorders of epithelial cell proliferation. The nucleic acids are also
CC   useful for chromosome identification, radiation hybrid mapping or long-
CC   range restriction mapping, as molecular weight markers, or as
CC   hybridization or diagnostic probes. The polypeptides and antibodies are
CC   useful for providing immunological probes for differential identification
CC   of the tissues immunohistochemistry assays. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 163 AA;

  Query Match             49.9%;  Score 762.5;  DB 6;  Length 163;
  Best Local Similarity   94.2%;  Pred. No. 3.2e-74;
  Matches  147;  Conservative    2;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              |||||||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGEXKCVVRVXECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQ 156
              |||||||||||||||||||||||    |: :||||
Db        121 CGWGKPISESLESVRLACIHTSPLIVSIYV-ELLRQ 155


RESULT 5
ADD37978
ID   ADD37978 standard; protein; 163 AA.
XX
AC   ADD37978;
XX
```

```
DT   15-JAN-2004  (first entry)
XX
DE   Human secreted protein #161.
XX
KW   human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
KW   Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.
XX
OS   Homo sapiens.
XX
PN   WO200290526-A2.
XX
PD   14-NOV-2002.
XX
PF   19-MAR-2002; 2002WO-US008279.
XX
PR   21-MAR-2001; 2001US-0277340P.
PR   19-JUL-2001; 2001US-0306171P.
PR   13-NOV-2001; 2001US-0331287P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-140218/13.
XX
PT   New human secreted proteins and nucleic acid molecules, useful for
PT   preparing a diagnostic or pharmaceutical composition for diagnosing or
PT   treating allergic or asthmatic disorders, or related immediate
PT   hypersensitivity disorders.
XX
PS   Claim 1; SEQ ID NO 460; 1323pp; English.
XX
CC   The present invention relates to an isolated polypeptide or human
CC   secreted protein. The polypeptides, nucleic acid molecules, antibodies or
CC   their fragments, and agonists or antagonists that bind are useful for
CC   preparing a diagnostic or pharmaceutical composition for diagnosing or
CC   treating allergic or asthmatic disorders. The polypeptide is also useful
CC   for identifying a binding partner by contacting the polypeptide with a
CC   binding partner, and determining whether the binding partner increases or
CC   decreases the activity of the polypeptide. The polypeptides and nucleic
CC   acid molecules are also useful for detecting, preventing, diagnosing,
CC   prognosticating, treating or ameliorating inflammatory disorders
CC   neoplastic diseases , wound healing and disorders of epithelial cell
CC   proliferation, immune disorders , cardiovascular disorders, blood-related
CC   disorders, infectious diseases, endocrine disorders , or gastrointestinal
CC   disorders. The nucleic acids are also useful for chromosome
CC   identification, radiation hybrid mapping or long-range restriction
CC   mapping, as molecular weight markers, or as hybridization or diagnostic
CC   probes. The polypeptides and antibodies are useful for providing
CC   immunological probes for differential identification of the tissues
CC   immunohistochemistry assays. The present sequence represents a human
CC   secreted protein.
XX
SQ   Sequence 163 AA;

  Query Match           49.9%;  Score 762.5;  DB 7;  Length 163;
  Best Local Similarity   94.2%;  Pred. No. 3.2e-74;
  Matches  147;  Conservative    2;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              |||||||||||||||||||||||||||||||||||||||| ||||||| |||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGEXKCVVRVXECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQ 156
              |||||||||||||||||||||    |: :||||
Db        121 CGWGKPISESLESVRLACIHTSPLIVSIYV-ELLRQ 155
```

RESULT 6

```
ABP62066
ID   ABP62066 standard; protein; 124 AA.
XX
AC   ABP62066;
XX
DT   12-NOV-2002  (first entry)
XX
DE   Human secreted protein SEQ ID NO 119.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200257420-A2.
XX
PD   25-JUL-2002.
XX
PF   17-JAN-2002; 2002WO-US001109.
XX
PR   18-JAN-2001; 2001US-0262066P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Moore PA,  Ruben SM,  Lafleur DW,  Shi Y,  Rosen CA,  Olsen H;
PI   Ebner R,  Brewer LA;
XX
DR   WPI; 2002-599716/64.
XX
PT   New polynucleotides and polypeptides useful for diagnosing, prognosing,
PT   treating or preventing e.g. neurodegenerative, central nervous system,
PT   autoimmune, respiratory, reproductive, or inflammatory diseases or
PT   disorders.
XX
PS   Claim 11; Page 754-755; 785pp; English.
XX
CC   The invention relates to novel genes (ABQ92553-ABQ92607) and proteins
CC   (ABP62013-ABP62153) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections
XX
SQ   Sequence 124 AA;

  Query Match           38.8%;  Score 593;  DB 5;  Length 124;
  Best Local Similarity   98.3%;  Pred. No. 6.1e-56;
  Matches  113;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
              |||| |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPSETEDVSNRNXVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
```

RESULT 7

```
ADA57562
ID   ADA57562 standard; protein; 124 AA.
XX
AC   ADA57562;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein #512.
XX
KW   immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
KW   cytostatic; cerebroprotective; neuroprotective; nootropic;
KW   cardiovascular; antiarteriosclerotic; gene therapy;
KW   human secreted protein; immune disorder; inflammation;
KW   respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
KW   inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
KW   multiple sclerosis; ischaemic brain injury; Parkinson's disease;
KW   Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
KW   triple helix formation; antisense gene therapy; forensic biology.
XX
OS   Homo sapiens.
XX
PN   WO2002102994-A2.
XX
PD   27-DEC-2002.
XX
PF   19-MAR-2002; 2002WO-US008278.
XX
PR   21-MAR-2001; 2001US-0277340P.
PR   19-JUL-2001; 2001US-0306171P.
PR   13-NOV-2001; 2001US-0331287P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-167512/16.
DR   N-PSDB; ADA56669.
XX
PT   New human secreted polypeptides and polynucleotides, useful for
PT   diagnosing, treating or preventing e.g. immune disorders, inflammatory
PT   conditions, respiratory disorders, cancers, CNS disorders, or
PT   neurodegenerative disorders.
XX
PS   Claim 13; SEQ ID NO 1755; 1754pp; English.
XX
CC   The invention relates to 592 new human secreted polypeptides useful for
CC   diagnosing, treating or preventing e.g. immune disorders, inflammatory
CC   conditions, respiratory disorders, cancers, CNS disorders, or
CC   neurodegenerative disorders, or polypeptides comprising an amino acid
CC   sequence at least 95% identical to the new sequences. The polypeptides,
CC   antibodies or antibody fragments that bind to the polypeptides, nucleic
CC   acids encoding the polypeptides, agonists or antagonists that binds to
CC   the polypeptide, are useful in preparing diagnostic or pharmaceutical
CC   compositions for diagnosing, treating or preventing an e.g. immune
CC   disorders, inflammatory conditions (e.g. inflammatory bowel disease,
CC   nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
CC   allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
CC   (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
CC   disorders (e.g. Parkinson's disease or Alzheimer's disease), and
CC   cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
CC   polynucleotides are useful for chromosome identification, chromosome
CC   mapping, for controlling gene expression through triple helix formation
CC   or antisense DNA or RNA, in gene therapy, for identifying individuals
CC   from minute biological samples, in forensic biology, and as hybridization
CC   probes. The polypeptides are useful for as molecular weight markers on
CC   sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
CC   gels, to raise antibodies, for testing biological activities, and for
CC   treating or preventing neural disorders, immune system disorders,
CC   muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC   renal, proliferative and/or cancerous diseases. This sequence corresponds
CC   to one of the polypeptide of the invention. Note: The sequence data for
CC   this patent did form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 124 AA;

  Query Match             38.8%;  Score 593;  DB 6;  Length 124;
  Best Local Similarity   98.3%;  Pred. No. 6.1e-56;
  Matches  113;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
              |||| |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPSETEDVSNRNXVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115


RESULT 8
ADA41456
ID   ADA41456 standard; protein; 124 AA.
XX
AC   ADA41456;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein.
XX
KW   Human; secreted protein; cancer; hyperproliferative disorder;
KW   rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW   anaemia; allergic reaction; asthma; cardiovascular disorder;
KW   wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective;
KW   antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW   vulnerary; cardiant; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO2002102993-A2.
XX
PD   27-DEC-2002.
XX
PF   19-MAR-2002; 2002WO-US008123.
XX
PR   21-MAR-2001; 2001US-0277340P.
PR   19-JUL-2001; 2001US-0306171P.
PR   13-NOV-2001; 2001US-0331287P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-175238/17.
XX
PT   New human secreted proteins and nucleic acid molecules, useful for
PT   preparing a diagnostic or pharmaceutical composition for diagnosing,
PT   preventing or treating cancer or other hyperproliferative disorder,
PT   asthma, allergies or AIDS.
XX
PS   Claim 1; SEQ ID NO 1839; 3205pp; English.
XX
CC   The invention relates to novel genes ADA39629-ADA40565 and proteins
CC   ADA40566-ADA41501 for human secreted proteins, useful for preventing,
CC   treating or ameliorating medical conditions e.g. by protein or gene
CC   therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC   fragments, and agonists or antagonists that bind to the polypeptide are
CC   useful for preparing a diagnostic or pharmaceutical composition for
CC   diagnosing or treating cancer or other hyperproliferative disorder. The
CC   polypeptides and nucleic acid molecules are also useful for detecting,
CC   preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC   or other hyperproliferative disorders including neoplasms, autoimmune
CC   disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC   erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC   anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC   thrombocytopenia), allergic reactions including asthma or eczema,
```

```
CC   inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC   bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC   Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC   (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC   fungal or viral infections including HIV/AIDS), or wound healing and
CC   disorders of epithelial cell proliferation. The nucleic acids are also
CC   useful for chromosome identification, radiation hybrid mapping or long-
CC   range restriction mapping, as molecular weight markers, or as
CC   hybridization or diagnostic probes. The polypeptides and antibodies are
CC   useful for providing immunological probes for differential identification
CC   of the tissues immunohistochemistry assays. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 124 AA;

  Query Match             38.8%;  Score 593;  DB 6;  Length 124;
  Best Local Similarity   98.3%;  Pred. No. 6.1e-56;
  Matches  113;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
              |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPSETEDVSNRNXVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115


RESULT 9
ADD38087
ID   ADD38087 standard; protein; 124 AA.
XX
AC   ADD38087;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human secreted protein #270.
XX
KW   human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
KW   Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.
XX
OS   Homo sapiens.
XX
PN   WO200290526-A2.
XX
PD   14-NOV-2002.
XX
PF   19-MAR-2002; 2002WO-US008279.
XX
PR   21-MAR-2001; 2001US-0277340P.
PR   19-JUL-2001; 2001US-0306171P.
PR   13-NOV-2001; 2001US-0331287P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-140218/13.
XX
PT   New human secreted proteins and nucleic acid molecules, useful for
PT   preparing a diagnostic or pharmaceutical composition for diagnosing or
PT   treating allergic or asthmatic disorders, or related immediate
PT   hypersensitivity disorders.
XX
PS   Claim 1; SEQ ID NO 569; 1323pp; English.
XX
CC   The present invention relates to an isolated polypeptide or human
CC   secreted protein. The polypeptides, nucleic acid molecules, antibodies or
CC   their fragments, and agonists or antagonists that bind are useful for
CC   preparing a diagnostic or pharmaceutical composition for diagnosing or
CC   treating allergic or asthmatic disorders. The polypeptide is also useful
CC   for identifying a binding partner by contacting the polypeptide with a
CC   binding partner, and determining whether the binding partner increases or
CC   decreases the activity of the polypeptide. The polypeptides and nucleic
CC   acid molecules are also useful for detecting, preventing, diagnosing,
CC   prognosticating, treating or ameliorating inflammatory disorders
CC   neoplastic diseases , wound healing and disorders of epithelial cell
CC   proliferation, immune disorders , cardiovascular disorders, blood-related
CC   disorders, infectious diseases, endocrine disorders , or gastrointestinal
CC   disorders. The nucleic acids are also useful for chromosome
CC   identification, radiation hybrid mapping or long-range restriction
CC   mapping, as molecular weight markers, or as hybridization or diagnostic
CC   probes. The polypeptides and antibodies are useful for providing
CC   immunological probes for differential identification of the tissues
CC   immunohistochemistry assays. The present sequence represents a human
CC   secreted protein.
XX
SQ   Sequence 124 AA;

  Query Match             38.8%;  Score 593;  DB 7;  Length 124;
  Best Local Similarity   98.3%;  Pred. No. 6.1e-56;
  Matches  113;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
              |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPSETEDVSNRNXVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115


RESULT 10
AAY12953
ID   AAY12953 standard; protein; 125 AA.
XX
AC   AAY12953;
XX
DT   17-JUN-1999  (first entry)
XX
DE   Amino acid sequence of a human secreted peptide.
XX
KW   Human secreted protein; cancer; immune disorder; infection;
KW   inflammatory disorder; skin disorder; tumour; atherosclerosis;
KW   restenosis; autoimmune disorder; Alzheimer's disease;
KW   peripheral neuropathy; trauma; spinal cord injury; allergy;
KW   hematopoietic disorder; skeletal disorder; neurological disorder;
KW   arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW   transplant rejection; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .28
FT                   /note= "signal peptide"
FT   Protein         29. .124
FT                   /note= "secreted protein"
XX
PN   WO9911293-A1.
XX
PD   11-MAR-1999.
XX
PF   03-SEP-1998;   98WO-US018360.
XX
PR   05-SEP-1997;   97US-0057626P.
PR   05-SEP-1997;   97US-0057663P.
PR   05-SEP-1997;   97US-0057669P.
PR   12-SEP-1997;   97US-0058666P.
PR   12-SEP-1997;   97US-0058667P.
PR   12-SEP-1997;   97US-0058973P.
PR   12-SEP-1997;   97US-0058974P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
```

PI Moore PA, Ruben SM, Lafleur DW, Shi Y, Rosen GA, Olsen HS;
PI Ebner R, Brewer LA;
XX
DR WPI; 1999-204988/17.
DR N-PSDB; AAX51740.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. neurological disorders,
PT tumors, immune disorders, inflammation or hematological disorders.
XX
PS Claim 11; Page 199; 215pp; English.
XX
CC AAY12914-68 represent human secreted proteins. The polypeptides and their
CC corresponding polynucleotides are useful for preventing, treating or
CC ameliorating medical conditions, e.g. by protein or gene therapy.
CC Pathological conditions can also be diagnosed by determining the amount
CC of the new polypeptides in a sample or by determining the presence of
CC mutations in the new polynucleotides. Specific uses are described for
CC each polynucleotide, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, immune disorders, infection, inflammatory disorders,
CC skin disorders, tumours, atherosclerosis, restenosis, autoimmune
CC disorders, Alzheimer's disease, peripheral neuropathies, trauma, spinal
CC cord injuries, allergy, hematopoietic disorders, skeletal disorders,
CC neurological disorders, arthritic disorders, asthma, immunodeficiency
CC diseases, AIDS and transplant rejection. The polypeptides are also useful
CC for identifying their binding partners
XX
SQ Sequence 125 AA;

Query Match 38.8%; Score 593; DB 2; Length 125;
Best Local Similarity 98.3%; Pred. No. 6.2e-56;
Matches 113; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
              |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPSETEDVSNRNXVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEEC 115
```

RESULT 11
ADN20693
ID ADN20693 standard; protein; 1315 AA.
XX
AC ADN20693;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #3346.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 3346; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought toleranoe,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1315 AA;

Query Match 7.2%; Score 110; DB 8; Length 1315;
Best Local Similarity 25.2%; Pred. No. 0.085;
Matches 61; Conservative 38; Mismatches 93; Indels 50; Gaps 14;

```
Qy         41 LAHEGEGEEETENNDSETAENYA---PPETEDV---SNRNVVKEVEFGMCTVTCGIGVRE 94
              |||| :||| :  |     : :||  |::|    |
Db        378 LAHESAANKETEEQEVPAKEEMLLLDHVDDDDVGEGSSKNAKTSKESN---------- 425

Qy         95 VILTNGCPGGES-KCVVRVEECRGPTDCGW----GKPISESL---ESVRLACIHTSP--- 143
                 ||  |  | :  :| ::  | || . |      |:|::|   | |||      |
Db        426 ---NNGAEGSSSGETFIRTDQLDGETD--WKLRLASPLSQNLSTEELVRLRVTAGKPDKR 480

Qy        144 LNRFKYMWKLL--RQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA----TIKP 197
              :| |   :|| ||| |    |    :|:  : ||: : :|   | ::|    |:
Db        481 VNEFVGTLELLPSRQDVMSGAAYNPREGDDVK--AAPLSID-NTAWANTVIASNATTLAV 537

Qy        198 TVYTSSELQMRRSSLPATDAALIF---VLTIGVIICVFIIFLLIFIIINWAAVKAFWGAK 254
              ||| : :   |: |:     :   : ::  |:|    || :|: |   |:: |  ||
Db        538 IVYTGPQTRSALSTSPSRSKTGLLEYEINSLTKILC----FLTLFLSIVLVALEGFSTAK 593

Qy        255 AS 256
              :
Db        594 GN 595
```

RESULT 12
AAY31833
ID AAY31833 standard; protein; 255 AA.
XX
AC AAY31833;
XX
DT 06-DEC-1999 (first entry)

```
XX
DE   Human foetal kidney secreted protein nn310_1.
XX
KW   Secreted protein; nn310_1; human; therapy; diagnosis; vaccine; kidney.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          20
FT                   /note= "a putative transmembrane domain is centered
FT                   around this residue"
FT   Domain          90
FT                   /note= "a putative transmembrane domain is centered
FT                   around this residue"
FT   Domain          230
FT                   /note= "a putative transmembrane domain is centered
FT                   around this residue"
XX
PN   WO9947555-A1.
XX
PD   23-SEP-1999.
XX
PF   18-MAR-1999;   99WO-US005939.
XX
PR   20-MAR-1998;   98US-0078803P.
PR   17-MAR-1999;   99US-00078803.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI   Merberg D, Treacy M, Agostino MJ, Steininger RJ;
XX
DR   WPI; 1999-562059/47.
DR   N-PSDB; AAZ19897.
XX
PT   New polynucleotides derived from murine fetal cell cDNA libraries,
PT   potentially used as, e.g. vaccines.
XX
PS   Claim 19(a); Page 98; 107pp; English.
XX
CC   This is the predicted amino acid sequence of a novel human secreted
CC   protein, nn310_1, as deduced from an isolated foetal kidney (293 cell)
CC   clone (see AAZ19897). The invention provides new human secreted proteins
CC   (see AAY31828-38) and polynucleotides (see AAZ19893-901) isolated from
CC   foetal cell, adult blood, adult brain and foetal kidney cDNA libraries.
CC   They are predicted to have biological activities which would make them
CC   suitable for treating, preventing or ameliorating medical conditions in
CC   humans and animals, although no supporting data are given. Suggested
CC   activities include nutritional, cytokine, tissue growth, cell
CC   proliferation and differentiation, immunostimulant (e.g. as vaccine),
CC   immunosuppressive, haematopoiesis regulating, activin or inhibin,
CC   chemotactic or chemokinetic, haemostatic or thrombolytic, receptor/ligand
CC   activity, antiinflammatory, cadherin or tumour invasion suppressor, and
CC   tumour inhibition activities
XX
SQ   Sequence 255 AA;

  Query Match             6.7%;  Score 103;  DB 2;  Length 255;
  Best Local Similarity   24.8%;  Pred. No. 0.042;
  Matches   41;  Conservative   21;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              |  :  | ||     :   |:  :    ||||||:| : | :   |  |   ||  |  ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTRRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                 ||         ||          |:|: :  |      |:: |:| |     :|  :
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE----------CDT--LDNNEIVATIKF 197
                 |      :  || : |:           ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKPKYAQDYDSGTYRCDVQLVKNLRLVKRLYF 176
```

```
RESULT 13
ABG02357
ID   ABG02357 standard; protein; 255 AA.
XX
AC   ABG02357;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #2348.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS66544.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 32716; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 255 AA;

  Query Match             6.7%;  Score 103;  DB 4;  Length 255;
  Best Local Similarity   24.8%;  Pred. No. 0.042;
  Matches   41;  Conservative   21;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              |  :  | ||     :   |:  :    ||||||:| : | :   |  |   ||  |  ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTRRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                 ||         ||          |:|: :  |      |:: |:| |     :|  :
```

```
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
              |       : || : |:           ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 14
AAU96170
ID   AAU96170 standard; protein; 255 AA.
XX
AC   AAU96170;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Human secreted protein, SEQ ID No 72.
XX
KW   Human; secreted protein; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; breast; liver; ischaemia;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   corneal infection; wound healing; ocular disorder; skin aging; sunburn;
KW   epithelial cell proliferation; organ transplantation; food additive;
KW   food storage.
XX
OS   Homo sapiens.
XX
PN   WO200224721-A1.
XX
PD   28-MAR-2002.
XX
PF   09-JAN-2001; 2001WO-US000544.
XX
PR   20-SEP-2000; 2000US-0234211P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Komatsoulis GA,  Baker KP,  Rosen CA,  Birse CE,  Soppet DR;
PI   Olsen HS,  Ni J,  Fiscella M,  Moore PA,  Wei P,  Ebner R,  Duan DR;
PI   Shi Y,  Choi GH;
XX
DR   WPI; 2002-330012/36.
DR   N-PSDB; ABK69091.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is used
PT   in preventing, treating or ameliorating a medical condition.
XX
PS   Claim 11; Page 498-499; 562pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) encoding a
CC   human secreted protein (II). (I) and (II) are used to prevent, treat or
CC   ameliorate a medical condition in e.g. humans, mice, rabbits, goats,
CC   horses, cats, dogs, chickens or sheep. (I) and (II) are also used in
CC   diagnosing a pathological condition or susceptibility to a pathological
CC   condition. The antibodies to (II) can also be used in alleviating
CC   symptoms associated with the disorders and in diagnostic immunoassays
CC   e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC   Disorders which are diagnosed or treated include autoimmune diseases e.g.
CC   rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the
CC   breast or liver, cardiovascular disorders e.g. cardiac arrest,
CC   cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
CC   system disorders e.g. Alzheimer's disease, infections caused by bacteria,
CC   viruses and fungi and ocular disorders e.g. corneal infection. The
CC   polypeptides can also be used to aid wound healing and epithelial cell
CC   proliferation, to prevent skin aging due to sunburn, to maintain organs
CC   before transplantation, for supporting cell culture of primary tissues,
CC   to regenerate tissues and in chemotaxis. The polypeptides can also be
CC   used as a food additive or preservative to increase or decrease storage
CC   capabilities. AAU96165-AAU96237 represent human secreted protein
CC   sequences and related sequences used in expression of the secreted
CC   proteins as described in examples of the invention
XX
SQ   Sequence 255 AA;

  Query Match             6.7%;  Score 103;  DB 5;  Length 255;
  Best Local Similarity   24.8%;  Pred. No. 0.042;
  Matches   41;  Conservative   21;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              | :  | ||    :   |: :    ||||||:| : | :   |  |   ||  |  ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTRRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
              ||          ||          .|:|: : .|       |:: |:| |     :|  :
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
              |       : || : |:           ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 15
ABG64819
ID   ABG64819 standard; protein; 255 AA.
XX
AC   ABG64819;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human albumin fusion protein #1494.
XX
KW   Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW   human serum albumin; HSA; cancer; reproductive disorder;
KW   digestive disorder; immune disorder; endocrine disorder;
KW   haematopoietic disorder; neural disorder; connective disorder;
KW   cytostatic; antiinfertility; antiinflammatory; antiulcer;
KW   immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
KW   neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW   osteopathic; antiarthritic.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200177137-A1.
XX
PD   18-OCT-2001.
XX
PF   12-APR-2001; 2001WO-US011988.
XX
PR   12-APR-2000; 2000US-0229358P.
PR   25-APR-2000; 2000US-0199384P.
PR   21-DEC-2000; 2000US-0256931P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Haseltine WA;
XX
DR   WPI; 2002-010886/01.
XX
PT   New fusion protein for treating disease e.g. diabetes comprises an
PT   albumin fused to a therapeutic protein.
XX
PS   Claim 1; Page 1522-1523; 2102pp; English.
XX
CC   The present invention relates to albumin fusion proteins comprising a
CC   therapeutic protein X and human albumin (HA, also known as human serum
CC   albumin, HSA). The proteins are useful for treating a disease or disorder
CC   that may be modulated by therapeutic protein X. The albumin extends the
CC   shelf-life of protein X, and may increase its biological in vitro/in vivo
CC   activity. The protein is useful for treating and diagnosing disorders
CC   such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
CC   disease, ulcerative colitis), immune disorders (e.g. acquired
CC   immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
CC   haematopoietic disorders, neural disorders (e.g. Alzheimer's,
CC   Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
CC   schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).
```

```
CC   ABG63326-ABG65518 represent albumin fusion proteins of the invention
XX
SQ   Sequence 255 AA;

  Query Match             6.7%;  Score 103;  DB 5;  Length 255;
  Best Local Similarity   24.8%;  Pred. No. 0.042;
  Matches   41;  Conservative   21;  Mismatches   65;  Indels   38;  Gaps    8;

Qy         58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
              | :  | ||    :   |:  :    ||||||:| : | :   |  |   ||  | ||
Db         25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTRRLECL 82

Qy        117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                 ||         ||           |:|: :  |       |:: |:| |      :|  :
Db         83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131

Qy        166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
              |       :  || : |:           ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


Search completed: April 15, 2005, 14:11:58
Job time : 175 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 15, 2005, 14:06:56 ; Search time 44 Seconds
(without alignments)
642.903 Million cell updates/sec

Title: US-10-809-655-9
Perfect score: 1528
Sequence: 1 MSPRGTGCSAGLLMTVGWLL.....LDQLPTEMPGEDDALSEWNE 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	97	6.3	709	2	S51793	dibasic processing
2	92.5	6.1	797	1	I46044	furin (EC 3.4.21.7
3	89.5	5.9	2181	2	A38198	calcium channel al
4	87.5	5.7	793	1	KXRTF	furin (EC 3.4.21.7
5	86.5	5.7	271	2	S12783	OX40 antigen precu
6	85.5	5.6	367	2	A32068	circumsporozoite p
7	85	5.6	378	1	OZZQAL	circumsporozoite p
8	84.5	5.5	387	2	C41156	circumsporozoite p
9	84	5.5	395	2	A41156	circumsporozoite p
10	84	5.5	594	2	I48771	Slp(w7) - mouse (f
11	83.5	5.5	610	2	T09988	probable transcrip
12	83	5.4	375	2	D75591	probable cation tr
13	83	5.4	496	2	E70142	glu-tRNA amidotran
14	83	5.4	970	2	S77349	excinuclease ABC c
15	82.5	5.4	429	2	A54504	circumsporozoite p
16	82.5	5.4	485	2	A60610	circumsporozoite p
17	82.5	5.4	573	2	T27671	hypothetical prote
18	82.5	5.4	999	2	T27628	hypothetical prote
19	82	5.4	228	2	T22924	hypothetical prote
20	82	5.4	530	2	JN0597	calnexin-like prot
21	82	5.4	1422	2	T42636	protein-tyrosine-p
22	81.5	5.3	378	1	OZZQAB	circumsporozoite p
23	81.5	5.3	611	2	F75095	probable asparagin
24	81.5	5.3	901	2	S65161	hypothetical prote
25	81.5	5.3	1238	2	T40120	C2H2 type zinc fin
26	81	5.3	387	2	D41156	circumsporozoite p
27	81	5.3	401	1	OZZQAC	circumsporozoite p
28	81	5.3	1232	2	T31426	cellulase (EC 3.2.
29	80.5	5.3	532	2	T49873	calnexin homolog -
30	80	5.2	315	2	T21907	hypothetical prote
31	80	5.2	522	2	T29705	hypothetical prote
32	80	5.2	794	1	KXHUF	furin (EC 3.4.21.7
33	80	5.2	1125	2	B41206	microtubule-associ
34	80	5.2	2203	2	T42742	voltage-dependent
35	79.5	5.2	363	1	OZZQAK	circumsporozoite p
36	79.5	5.2	597	2	G83990	two-component sens
37	79.5	5.2	793	1	KXMSF	furin (EC 3.4.21.7
38	79	5.2	158	2	H87711	hypothetical prote
39	79	5.2	545	2	T47952	hypothetical prote
40	79	5.2	795	2	T20939	hypothetical prote
41	78.5	5.1	380	2	T25819	hypothetical prote
42	78.5	5.1	569	2	T22928	hypothetical prote
43	78.5	5.1	1061	2	F86211	hypothetical prote
44	78	5.1	518	2	C89785	hypothetical prote
45	78	5.1	621	2	H84922	hypothetical prote

ALIGNMENTS

RESULT 1
S51793
dibasic processing endoproteinase precursor (EC 3.4.21.-) - fission yeast (Schizosaccharc
C;Species: Schizosaccharomyces pombe
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S51793; T38166
R;Davey, J.; Davis, K.; Imai, Y.; Yamamoto, M.; Matthews, G.
EMBO J. 13, 5910-5921, 1994
A;Title: Isolation and characterization of krp, a dibasic endopeptidase required for cell
A;Reference number: S51793; MUID:95112801; PMID:7813430
A;Accession: S51793
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-709 <DAV>
A;Cross-references: UNIPROT:Q09175; EMBL:X82435; NID:g565060; PIDN:CAA57818.1; PID:g5650
R;Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z21775
A;Accession: T38166
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-709 <DEV>
A;Cross-references: EMBL:Z70043; NID:g1220275; PIDN:CAA93896.1; GSPDB:GN00066; SPDB:SPAC
A;Experimental source: strain 972h-; cosmid c22E12
C;Genetics:
A;Gene: SPAC22E12.09c
A;Map position: 1
C;Superfamily: Saccharomyces kexin; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;153-385/Domain: subtilisin homology <SBT>
F;162,200,371/Active site: Asp, His, Ser #status predicted

Query Match 6.3%; Score 97; DB 2; Length 709;
Best Local Similarity 25.8%; Pred. No. 2.8;
Matches 40; Conservative 17; Mismatches 52; Indels 46; Gaps 7;

```
Qy        150 MWKLLRQDQQ----------SIILVNDSAILEVRKESHPLAFECDTLD---------- 187
              :||||  |:           : |  :|   |    : || :  |||:
Db        558 VWKLLVNDRSGGKHEGTFENWQLALWGES---ENPSNTAPLPY--DTLELPKEMVLGIYS 612

Qy        188 --NNEIV--------ATIKFTVYTSSELQMRRSS--------LPATDAAL-IFVLTIGVI 228
                |:::          :  || || |       |:        || |   |      |
Db        613 EPNSDLTNSSTLLSPTSTSFTSYTVSATATPTSTSHIPIPTVLPPTQPVLEPSYREIVAF 672

Qy        229 ICVFIIFLLIPIIINWAAVKAFWGAKASTPEVQSE 263
              |  |::|  ||: : |  : ||| |||  |  | |
Db        673 ITFFLLFAFIFVAVIWTWISAFWKAKAPPPLSQQE 707
```

RESULT 2
I46044
furin (EC 3.4.21.75) precursor - bovine

```
N;Alternate names: kexin homolog; paired-basic endopeptidase; prohormone-processing endo
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: I46044; S41191
R;Vey, M.; Schafer, W.; Berghofer, S.; Klenk, H.D.; Garten, W.
J. Cell Biol. 127, 1829-1842, 1994
A;Title: Maturation of the trans-Golgi network protease furin: compartmentalization of p
A;Reference number: A55189; MUID:95105228; PMID:7806563
A;Accession: I46044
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-797 <VEY>
A;Cross-references: UNIPROT:Q28193; EMBL:X75956; NID:g439648; PIDN:CAA53569.1; PID:g4396
C;Function:
A;Description: cleavage of precursor proteins during constitutive secretory pathway at R
C;Superfamily: kexin; subtilisin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-797/Product: furin #status predicted <MAT>
F;27-718/Domain: extracellular #status predicted <EXT>
F;144-382/Domain: subtilisin homology <SBT>
F;719-739/Domain: transmembrane #status predicted <TMM>
F;740-797/Domain: intracellular #status predicted <INT>
F;153,194,368/Active site: Asp, His, Ser #status predicted
F;387,440,553/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             6.1%;  Score 92.5;  DB 1;  Length 797;
  Best Local Similarity   21.7%;  Pred. No. 7.8;
  Matches   76;  Conservative   32;  Mismatches  103;  Indels  139;  Gaps   18;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGL---QSARGTNVTAAVQDAGLAHEGEGEEETENNDSE 57
              :|| ||  :           |||      || | |  :       :  ||    |  ::
Db        506 VSPMGTRST----------LLAARPHDYSADGFNDWAFMTTHSWDEDPSGEWVLEIENTS 555

Qy         58 TAENY------------------APPE-----TEDVSNRNVVKEVEFGM----CTVTCG 89
               | ||                  |||      |   |   || |  | :    |   |
Db        556 EANNYGTLTKFTLVLYGTAPEGLPTPPESIGCKTLTSSQACVVCEEGFSLHQKNCVQHCP 615

Qy         90 IGVR--------------EVILTNGCPGGESKCVVRVEECRG--PTDCGWGKPISESLES 133
               |                |:|  : |    : |      |:|  ||||    |   ||:
Db        616 PGFAPQVLDTHYSTENDVEIIRASVCTPCHASCAT----CQGPAPTDC-LSCPSHASLDP 670

Qy        134 VRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESH---------PLAFECD 184
              |   |              || | |             :|||           |   |
Db        671 VEQTC--------------SRQSQSS------------RESHQQQPPPPPRPPPAE-- 700

Qy        185 TLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFIIINW 244
              |||          | ::|   ||:    ::  |:   |:   ::|: :|:::
Db        701 -------VAT---------EPRLRADLLPSHLPEVVAGLSCAFIV---LVFVTVFLVLQL 741

Qy        245 AAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
               :  :| | | |     ::  : ||       || |         | |
Db        742 RSGFSFRGVKVYT----MDRGLISYKG------LPPE---------AWQE 772


RESULT 3
A38198
calcium channel alpha-1 chain, pancreatic - human
N;Alternate names: beta-cell-type calcium channel alpha-1 chain; neuroendocrine-type cal
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000
C;Accession: A38198
R;Seino, S.; Chen, L.; Seino, M.; Blondel, O.; Takeda, J.; Johnson, J.H.; Bell, G.I.
Proc. Natl. Acad. Sci. U.S.A. 89, 584-588, 1992
A;Title: Cloning of the alpha 1 subunit of a voltage-dependent calcium channel expressed
A;Reference number: A38198; MUID:92115705; PMID:1309948
A;Accession: A38198
A;Molecule type: mRNA
A;Residues: 1-2181 <SEI>
A;Cross-references: GB:M83566; NID:g179751; PIDN:AAA35629.1; PID:g179752
A;Experimental source: pancreatic beta cells
C;Superfamily: voltage-dependent calcium channel protein alpha-1 chain
C;Keywords: membrane protein; voltage-gated ion channel

  Query Match             5.9%;  Score 89.5;  DB 2;  Length 2181;
  Best Local Similarity   21.2%;  Pred. No. 43;
  Matches   71;  Conservative   46;  Mismatches   97;  Indels  121;  Gaps   18;

Qy         32 VTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIG 91
              :| | :|    :| || || : | |          ||| |:  ||  | |
Db        442 ITQA-EDIDPENEEEGGEEGKRNTSMPTS-----ETESVNTENVSGEGE---------- 484

Qy         92 VREVILTNGCPG---------GESKCVVRVEECRGPTDC-GWGKPISESLESVR------ 135
                 || |          | :|          ||: |  ||: ||:|  | |
Db        485 ------NRGCCGSLWCWWRRRGAAKA--------GPSGCRRWGQAISKSKLSRRWRRWNR 530

Qy        136 ---------------------LACIHTSPLNRFKYM---WKLLRQD-QQSIILVNDSAIL 170
                                  | ::|  ::    |    |    ||     ::|   :  :
Db        531 FNRRRCRAAVKSVTFYWLVIVLVFLNTLTISSEHYNQPDWLTQIQDIANKVLLALFTCEM 590

Qy        171 EVRKESHPL---------APEC---------DTLDNNEIVATIKFTVYTSSEL----QMR 208
              |:  |  |          |:|         |   ||:: :  :|:    |    ::
Db        591 LVKMYSLGLQAYFVSLFNRFDCFVVCGGITETILVELEIMSPLGISVFRCVRLLRIFKVT 650

Qy        209 R--SSLPATDAALI-FVLTIGVIICVFIIFLLIFIIINWAAVKAFWGAKASTPEVQSEQS 265
              |  :||    |:|:  : :|  :: :  :|::|| ::        :| | :  | |:::|
Db        651 RHWTSLSNLVASLLNSMKSIASLLLLLFLFIIIFSLLG----MQLFGGKFNFDETQTKRS 706

Qy        266 SVRYKDSTSLDQLPTE-------MPGEDDALSEWN 293
                : |  |             : |||     ||
Db        707 --------TFDNFPQALLTVFQILTGED-----WN 728


RESULT 4
KXRTF
furin (EC 3.4.21.75) precursor - rat
N;Alternate names: kexin homolog; paired-basic endopeptidase; prohormone-processing endo
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: S13106
R;Misumi, Y.; Sohda, M.; Ikehara, Y.
Nucleic Acids Res. 18, 6719, 1990
A;Title: Sequence of the cDNA encoding rat furin, a possible propeptide-processing endop
A;Reference number: S13106; MUID:91067492; PMID:2251148
A;Accession: S13106
A;Molecule type: mRNA
A;Residues: 1-793 <MIS>
A;Cross-references: UNIPROT:P23377; EMBL:X55660; NID:g56171; PIDN:CAA39193.1; PID:g56172
C;Comment: This subtilisin-like endoproteinase removes paired basic residues to process p
C;Superfamily: kexin; subtilisin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-793/Product: furin #status predicted <MAT>
F;27-714/Domain: extracellular #status predicted <EXT>
F;144-382/Domain: subtilisin homology <SBT>
F;715-735/Domain: transmembrane #status predicted <TMM>
F;736-793/Domain: intracellular #status predicted <INT>
F;153,194,368/Active site: Asp, His, Ser #status predicted
F;387,440,553/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             5.7%;  Score 87.5;  DB 1;  Length 793;
  Best Local Similarity   21.7%;  Pred. No. 21;
  Matches   75;  Conservative   30;  Mismatches  106;  Indels  135;  Gaps   17;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGL---QSARGTNVTAAVQDAGLAHEGEGEEETENNDSE 57
              :|| ||  :           |||      || | |  | :       :  ||    |  ::
Db        506 ISPMGTRST----------LLAARPHDYSADGFNDWAFMTTHSWDEDPSGEWVLEIENTS 555

Qy         58 TAENY------------------APPE-----TEDVSNRNVVKEVEFGM----CTVTCG 89
               | ||                  ||||     |   |   || |  | :    |   |
Db        556 EANNYGTLTKFTLVLYGTASEGLSAPPESSGCKTLTSSQACVVCEEGPSLHQKSCVQRCP 615

Qy         90 IGVR--------------EVILTNGCPGGESKCVVRVEECRG--PTDCGWGKPISESLES 133
                |              |:|  : |    : |      |:|  ||||    |   ||:
Db        616 PGFTPQVLDTHYSTENDVEIIRASVCTPCHASCAT----CQGPAPTDC-LSCPSHASLDP 670
```

```
Qy        134 VRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA 193
              |   |                 || | |          | |.:  | |   :
Db        671 VEQTC---------------SRQSQSS---------RESRPQQPPPALRPEV-------- 698

Qy        194 TIKFTVYTSSELQMR-----RSSLPATDAALIFVLTIGVIICVFIIFLLIFIIINWAAVK 248
                  |:: |      | ||   | |  ::      : :|| ::|: ::  :
Db        699 ----------EVEPRLRAGLASHLPEVLAGLSCLI-------IALIFGIVFLFLHRCSGF 741

Qy        249 AFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              :| | |  |     ::  : ||       || |          | |
Db        742 SFRGVKVYT----MDRGLISYKG------LPPE---------AWQE 768


RESULT 5
S12783
OX40 antigen precursor - rat
N;Alternate names: nerve growth factor receptor homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: S12783; S08036
R;Mallett, S.; Fossum, S.; Barclay, A.N.
EMBO J. 9, 1063-1068, 1990
A;Title: Characterization of the MRC OX40 antigen of activated CD4 positive T lymphocyte
A;Reference number: S12783; MUID:90214614; PMID:2157591
A;Accession: S12783
A;Molecule type: mRNA
A;Residues: 1-271 <MAL>
A;Cross-references: UNIPROT:P15725; EMBL:X17037; NID:g57830; PIDN:CAA34897.1; PID:g57831
C;Superfamily: CD27 antigen; NGF receptor repeat homology
C;Keywords: growth factor receptor; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-271/Product: OX40 antigen #status predicted <MAT>
F;211-235/Domain: transmembrane #status predicted <TMM>

  Query Match             5.7%;  Score 86.5;  DB 2;  Length 271;
  Best Local Similarity   21.1%;  Pred. No. 7.7;
  Matches   42;  Conservative   35;  Mismatches   69;  Indels   53;  Gaps   10;

Qy         99 NGCPGGESKCVVRVEECRGPTDCGWGKPISESLESVRLACI-----HTSP-LNRFKYMWK 152
              |  |  :: |     :||  |     :|  :|   : : |:     | ||  |:    |
Db         93 NCTPTEDTVC-----QCRPGT-----QPRQDSSHKLGVDCVPCPPGHFSPGSNQACKPW- 141

Qy        153 LLRQDQQSIILVNDSAILEVRKESHPLAFECDTL-DNNEIVATIKF---------TVYTS 202
                          :  |   ::  || :   ||: ::  ::||: :           |   |
Db        142 ------------TNCTLSGKQIRHPASNSLDTVCEDRSLLATLLWETQRTTFRPTTVPS 188

Qy        203 SELQMRRSSLPATDA---------ALIFVLTIGVIICVFIIFLLIFIIINWAA---VKAF 250
              : :  | | ||:|          :|  | :|::  : ::  |  :    | :    |
Db        189 TTVWPRTSQLPSTPTLVAPEGPAFAVILGLGLGLLAPLTVLLALYLLRKAWRSPNTPKPC 248

Qy        251 WGAKASTPEVQSEQSSVRY 269
              ||    || :| ||:   :
Db        249 WGNSFRTP-IQEEQTDTHF 266


RESULT 6
A32068
circumsporozoite protein - Plasmodium vivax (strain North Korean) (fragment)
C;Species: Plasmodium vivax
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 09-Jul-2004
C;Accession: A32068
R;Arnot, D.E.; Barnwell, J.W.; Stewart, M.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 8102-8106, 1988
A;Title: Does biased gene conversion influence polymorphism in the circumsporozoite prot
A;Reference number: A32068; MUID:89042133; PMID:3054880
A;Accession: A32068
A;Molecule type: DNA
A;Residues: 1-367 <ARN>
A;Cross-references: UNIPROT:Q26169; GB:M20670; GB:J04090; NID:g160183; PIDN:AAA29534.1;
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F;292-345/Domain: thrombospondin type 1 repeat homology <THR1>

  Query Match             5.6%;  Score 85.5;  DB 2;  Length 367;
  Best Local Similarity   30.5%;  Pred. No. 13;
  Matches   25;  Conservative    8;  Mismatches   28;  Indels   21;  Gaps    2;

Qy         22 AGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNVVKE--- 78
              || |:| |        :|||     |    : :||:   |              |  |||
Db        244 AGGQAAGGNAANKKAEDAGGNAGGNAGGQGQNNEGANA-----------PNEKSVKEYLD 292

Qy         79 -------VEFGMCTVTCGIGVR 93
                     |:  |:||||:|||
Db        293 KVRATVGTEWTPCSVTCGVGVR 314


RESULT 7
OZZQAL
circumsporozoite protein precursor - Plasmodium cynomolgi (strain London)
N;Alternate names: major sporozoite surface antigen
C;Species: Plasmodium cynomolgi
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C;Accession: A26255
R;Galinski, M.R.; Arnot, D.E.; Cochrane, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Enea, V
Cell 48, 311-319, 1987
A;Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A;Reference number: A90889; MUID:87102878; PMID:3802196
A;Accession: A26255
A;Molecule type: DNA
A;Residues: 1-378 <GAL>
C;Comment: There are three distinct regions in the mature circumsporozoite protein, the a
obic membrane-anchoring sequence.
C;Comment: There are 19 tandem copies of a 6-residue repeat and 6 copies of an 11-residue
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C;Keywords: sporozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-378/Product: circumsporozoite protein #status predicted <MAT>
F;98-211/Region: 6-residue repeats
F;212-277/Region: 11-residue repeats
F;303-356/Domain: thrombospondin type 1 repeat homology <THR1>

  Query Match             5.6%;  Score 85;  DB 1;  Length 378;
  Best Local Similarity   29.7%;  Pred. No. 15;
  Matches   38;  Conservative   11;  Mismatches   51;  Indels   28;  Gaps    4;

Qy         22 AGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNVVKE--- 78
              || |:  | |        || |  |:|    :||           |  :: |  :|||
Db        259 AGGQAGAGGNQAGGQAGAGDAGAGQG----QNN-----------EGANMPNAKLVKEYLD 303

Qy         79 -------VEFGMCTVTCGIGV---REVILTNGCPGGESKCVVRVEECRGPTDCGWGKPIS 128
                     ||:  |:|||| ||   |:|    |  |        :  | |       |    :|
Db        304 KIRSTIGVEWSPCSVTCGKGVRMRRKVNAANKKPEELDVNDLETEVCTMDKCAGIFNVVS 363

Qy        129 ESLESVRL 136
                ||  | |
Db        364 NSLGLVIL 371


RESULT 8
C41156
circumsporozoite protein - Plasmodium vivax (isolate B7-4)
C;Species: Plasmodium vivax
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 09-Jul-2004
C;Accession: C41156
R;Qari, S.H.; Goldman, I.F.; Povoa, M.M.; Oliveira, S.; Alpers, M.P.; Lal, A.A.
J. Biol. Chem. 266, 16297-16300, 1991
A;Title: Wide distribution of the variant form of the human malaria parasite Plasmodium v
A;Reference number: A41156; MUID:91358402; PMID:1885563
A;Accession: C41156
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-387 <QAR>
A;Cross-references: UNIPROT:Q7M3W8; GB:M69062
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F;312-365/Domain: thrombospondin type 1 repeat homology <THR1>

  Query Match             5.5%;  Score 84.5;  DB 2;  Length 387;
```

```
  Best Local Similarity   31.4%;  Pred. No. 17;
  Matches   38;  Conservative   15;  Mismatches   51;  Indels   17;  Gaps    5;

Qy         22 AGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNVVKE--- 78
              || |:| |    || : || |  |:|    :||:      || |   :   : |:
Db        271 AGGQAAGG---NAANKKAGDAGAGQG----QNNEGAN----APNEKSVIEYLDKVRATVG 319

Qy         79 VEFGMCTVTCGIGV---REVILTNGCPGGESKCVVRVEECRGPTDCGWGKPISESLESVR 135
               |:  |:||||:||   | |  ||  |   :  :  : |     |    :| ||  |
Db        320 TEWTPCSVTCGVGVRVRRRVNATNKKPEDLTLNDLETDVCTMDKCAGIFNVVSNSLGLVI 379

Qy        136 L 136
               |
Db        380 L 380


RESULT 9
A41156
circumsporozoite protein - Plasmodium vivax (isolate P19/D)
C;Species: Plasmodium vivax
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 09-Jul-2004
C;Accession: A41156
R;Qari, S.H.; Goldman, I.F.; Povoa, M.M.; Oliveira, S.; Alpers, M.P.; Lal, A.A.
J. Biol. Chem. 266, 16297-16300, 1991
A;Title: Wide distribution of the variant form of the human malaria parasite Plasmodium
A;Reference number: A41156; MUID:91358402; PMID:1885563
A;Accession: A41156
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-395 <QAR>
A;Cross-references: UNIPROT:Q7M3X0; GB:M69059
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F;320-373/Domain: thrombospondin type 1 repeat homology <THR1>

  Query Match             5.5%;  Score 84;  DB 2;  Length 395;
  Best Local Similarity   34.1%;  Pred. No. 19;
  Matches   28;  Conservative    9;  Mismatches   17;  Indels   28;  Gaps    4;

Qy         22 AGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNVVKE--- 78
              || |:| |    || : || |  |:|    :||:    |           :|   |||
Db        279 AGGQAAGG---NAANKKAGDAGAGQG----QNNEGANA----------TNEKSVKEYLD 320

Qy         79 -------VEFGMCTVTCGIGVR 93
                    |:  |:||||:|||
Db        321 KVRATVGTEWTPCSVTCGVGVR 342


RESULT 10
I48771
Slp(w7) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: I48771
R;Hemenway, C.; Kalff, M.; Stavenhagen, J.; Walthall, D.; Robins, D.
Nucleic Acids Res. 14, 2539-2554, 1986
A;Title: Sequence comparison of alleles of the fourth component of complement (C4) and s
A;Reference number: I48274; MUID:86176748; PMID:3008092
A;Accession: I48771
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-594 <RES>
A;Cross-references: UNIPROT:Q62238; EMBL:X06454; NID:g54105; PIDN:CAA29760.1; PID:g54106
C;Superfamily: alpha-2-macroglobulin

  Query Match             5.5%;  Score 84;  DB 2;  Length 594;
  Best Local Similarity   22.9%;  Pred. No. 30;
  Matches   50;  Conservative   31;  Mismatches   77;  Indels   60;  Gaps   11;

Qy         20 LLAGLQSARG-TNVTAAVQDAGLAH-EGEGEEETENNDSETAENYAPPETEDVSNRNVVK 77
              ||:|  : ||    .  :: |  ::| | :|       ||       | | :       :
Db        340 LLSGFHALRGDLEKLTSLSDRYVSHFETDGPHVLLYFDS-------VPTTRECVGFGASQ 392

Qy         78 EVEFGM----------------CTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
                 ||  |:                  |:|      :  :|   | |        : |:
Db        393 EVVVGLVQPSSAVLYDYYSPDHKCSVFYAAPTKSQLLATLCSG---------DVCQ---- 439

Qy        121 CGWGK--PISESLE---------SVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAI 169
              |  ||  |:  |||          :| || :  |   : :  |:||:| ::   : :| |
Db        440 CAQGKCPPLLRSLERRVEDKDGYRMRFACYY--PRVEYGFTVKVLREDGRAAFRLFESKI 497

Qy        170 LEVRKESHPLAFECDTLDNNEIVATIKFTVYTSSELQM 207
               :|      | |  ||: :  |  |  |    ||  |::
Db        498 TQV------LHFRTDTMAS--IGQTRNFLSRTSCRLRL 527


RESULT 11
T09988
probable transcription termination factor - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T09988
R;Robison, K.
submitted to the EMBL Data Library, September 1994
A;Reference number: Z16911
A;Accession: T09988
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-610 <ROB>
A;Cross-references: UNIPROT:P45835; EMBL:U15186; NID:g699323; PID:g699332
C;Keywords: ATP; transcription termination

  Query Match             5.5%;  Score 83.5;  DB 2;  Length 610;
  Best Local Similarity   24.1%;  Pred. No. 34;
  Matches   60;  Conservative   30;  Mismatches  102;  Indels   57;  Gaps   14;

Qy          4 RGTGCSAGLLMTVGWL---LLAGLQSARG-TNVTAAVQDAGLAHEGEGEE-ETENNDSET 58
              |      ||:  | |      |:| ::  ||  | |:          | |       ||   ::
Db         50 RALANQAGVKGTSGMRKSELIAAIEECRGQANGTSVNDGPSRDHGGSATAISTEALAAQE 109

Qy         59 AENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVIL----------TNGCPGGESKC 108
              :|||  |       |   :| :     || |     |                |    ||:|
Db        110 EQNYAIVEVSRRERRGASREAD-----VTAGTSTAEATESDCQGTADDDTRTLQGGQSD- 163

Qy        109 VVRVEECRGPTDCGWGKPISESLESVRLACIHTSPL------------NRFKYMWKLLR 155
               : || ||| | |  : : :   |::       |                ||:   :  |
Db        164 -TKTEE-RGP-DVGNDQGVEQQSSSLQ-------PRGDDDGEGRQGRRGRRFRDRDRRRR 213

Qy        156 QDQQSIILVNDSAILEVRKES--HPLAFECDTLDNNEIVATIKF-----TVYTSSELQMR 208
               ::   .   | |  |:|::    |:|   | |||   | |  :      || |  : :|
Db        214 GERS-----GDGAEAELRQDDVVQPVAGILDVLDNYAFVRTSGYLAGPHDVYVSMSM-VR 267

Qy        209 RSSLPATDA 217
              ::  |   ||
Db        268 KNGLRRGDA 276


RESULT 12
D75591
probable cation transporter - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: D75591
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mal
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: D75591
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-375 <WHI>
A;Cross-references: UNIPROT:Q9RYF7; GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12428
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0361
```

```
A;Map position: 2
C;Superfamily: magnesium and cobalt transport protein

  Query Match             5.4%;  Score 83;  DB 2;  Length 375;
  Best Local Similarity   24.7%;  Pred. No. 22;
  Matches   60;  Conservative   27;  Mismatches   78;  Indels   78;  Gaps   13;

Qy        116 RGPTDCGWGKPISE-SLESVRLACIHTSPLNRPKY------MWKLLRQDQQSII---LVN 165
              | ||| | | |: : | :    |   |     : ::      : : |  | |: :   |||
Db         24 RPPTDAGPGDPLLDGSGQPAPHAEAHVFRDGKVRHLPLDLNLARGLLDDPQAFVWFDLVN 83

Qy        166 -DSAILEVRKES---HPLAFECDTLDNNEIVA-----TIKFTVYTSSELQMRRSSLPATD 216
               | | ||  :|    ||||  | | |   : |      | | |       :  :|||: |
Db         84 PDPADLEALREPFDLHPLAIE-DALVGGQRVKVEPYDTFWFVV-------LHGASLPSQD 135

Qy        217 AALI--FVLTIGVIICVFIIFLLIF----IIINWAAVKAFWGAKAS-------------- 256
                :    | ||    : |    :|    |:  |  | | | : ||
Db        136 RLQLHEMALFIGQRFVLTIQHQPLFADEEIVERWQLVPAAWRSSASSLTYVILDTIVDNL 195

Qy        257 ---TPEVQSEQSSVR--------------------------YKDSTSL-DQLPTEMPGE 285
                | ::: |   ||                          | : || | ||| :
Db        196 RELTDQIERELREVRTTMTRSQVVHPELLQRIFSLEEVTHEAYTVALSLRDTLPTFVHAP 255

Qy        286 DDA 288
               ::|
Db        256 EEA 258


RESULT 13
E70142
glu-tRNA amidotransferase, subunit A (gluA) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: E70142
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: E70142
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-496 <KLE>
A;Cross-references: UNIPROT:O51317; GB:AE001140; GB:AE000783; NID:g2688233; PIDN:AAC6671
A;Experimental source: strain B31
C;Superfamily: indoleacetamide hydrolase

  Query Match             5.4%;  Score 83;  DB 2;  Length 496;
  Best Local Similarity   19.4%;  Pred. No. 30;
  Matches   43;  Conservative   44;  Mismatches   81;  Indels   54;  Gaps    8;

Qy         47 GEEETENNDSETAENYAPPETEDVSNRN--VVKEV-------------EFGMCTVTCGI 90
              | :: ::    :  ::: | : | :  :|  |:||:            :| :  :: ||
Db        246 GSDKMDSTSVDIFDDFYPLKIESLQGKNLAVIKELSEDLMDKNVANSFAKFKLDLLSKGI 305

Qy         91 GVREVILTNGCPGGESKCVVRVEECRGPTDCG----------WGKPISESLESVRLACIH 140
               ::|| :          |   :: :      | :           :|| ||| |
Db        306 NIKEVSIE------EINFILSIYYIISPVEASSNLARYTGLCYGKRISEGLS-------- 351

Qy        141 TSPLNRPKYMWK---LLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDN------NEI 191
                || | :  :    |  :: :: |:| |   :|  | :|    |  |: | |       |::
Db        352 ---LNDFYFKHRSNFLSEEVKRRIVLGN--YLLSERYDSKYYAKACEILQNLIIPKFNKL 406

Qy        192 VATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFI 233
                :  | :  :| ::  |  |    |   ::   |  :|   |
Db        407 FESCDFIITPTSFVKPFRLGLDFDDPVKMYYSDICTVIANLI 448


RESULT 14
877349
excinuclease ABC chain A - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein slr1844; uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S77349
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S77349
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-970 <KAN>
A;Cross-references: UNIPROT:P73412; EMBL:D90906; GB:AB001339; NID:g1652492; PIDN:BAA17452
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: uvrA
C;Function:
A;Description: has ATPase and DNA binding activity; involved in DNA repair
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; nucleotide binding; P-l
F;34-41/Region: nucleotide-binding motif A (P-loop)
F;652-935/Domain: ATP-binding cassette homology <ABCE>
F;669-676/Region: nucleotide-binding motif A (P-loop)

  Query Match             5.4%;  Score 83;  DB 2;  Length 970;
  Best Local Similarity   25.6%;  Pred. No. 63;
  Matches   42;  Conservative   21;  Mismatches   65;  Indels   36;  Gaps    6;

Qy         24 LQSARGTNV-------TAAVQDAGLAHEGEGEEETENNDSETA---------ENYAPPE- 66
              |: | || :       | || | |   : |  :  ||  :  |          ||:| ||
Db        228 LKQAEGTAIIDILDKPTLAVLDGGKKDDKEALKAAENGQAYHAELPKEIIFSENFACPEH 287

Qy         67 ---TEDVSNRNVVKEVEFGMCTVTCGIG-VREVILTNGCPGGESKCVVRVEECRGPTDCG 122
                 :::| |       :| |    ||| ||         |  |    | :
Db        288 GAVMDELSPRLFSFNSPYGACPDCHGIGFVRSFCPDLVIPDPEKPVYVAI---------- 337

Qy        123 WGKPISESLESVRLACIHTSPLN---RFKYMWKLLRQDQQSIIL 163
                | ||   |  |: :::   :   : :  || | ::|: |||
Db        338 --APWSEKDNSYYLSLLYSLGQHFDFQLQTPWKKLTKEQKEIIL 379


RESULT 15
A54504
circumsporozoite protein - Plasmodium malariae (strain Uganda 1-CDC)
C;Species: Plasmodium malariae
C;Date: 06-Oct-1994 #sequence_revision 04-Nov-1994 #text_change 09-Jul-2004
C;Accession: A54504
R;Lal, A.A.; de la Cruz, V.F.; Campbell, G.H.; Procell, P.M.; Collins, W.E.; McCutchan, T
Mol. Biochem. Parasitol. 30, 291-294, 1988
A;Title: Structure of the circumsporozoite gene of Plasmodium malariae.
A;Reference number: A54504; MUID:89040027; PMID:3054537
A;Accession: A54504
A;Molecule type: DNA
A;Residues: 1-429 <LAL>
A;Cross-references: UNIPROT:P13815; GB:J03992; NID:g160220; PIDN:AAA29557.1; PID:g160221
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C;Keywords: tandem repeat
F;354-407/Domain: thrombospondin type 1 repeat homology <THR1>

  Query Match             5.4%;  Score 82.5;  DB 2;  Length 429;
  Best Local Similarity   35.1%;  Pred. No. 28;
  Matches   20;  Conservative    9;  Mismatches   17;  Indels   11;  Gaps    3;

Qy         42 AHEGEGEEETENNDSETAENYAPPETEDVSN-----RNVVKEVEFGMCTVTCGIGVR 93
              |:  : : : |||||         |  | : |     || : | |:  |:|||| |:|
Db        326 ANTNKKDNQEENNDSSNG-----PSEEHIKNYLESIRNSITE-EWSPCSVTCGSGIR 376


Search completed: April 15, 2005, 14:16:01
```

Job time : 46 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 15, 2005, 14:04:30 ; Search time 187 Seconds
(without alignments)
805.087 Million cell updates/sec

Title: US-10-809-655-9
Perfect score: 1528
Sequence: 1 MSPRGTGCSAGLLMTVGWLL.....LDQLPTEMPGEDDALSEWNE 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1528	100.0	294	2	Q9HBV2	Q9hbv2 homo sapien
2	1115.5	73.0	305	2	Q9DA48	Q9da48 mus musculu
3	311	20.4	90	2	Q80VM2	Q80vm2 mus musculu
4	110	7.2	1331	2	Q9HED0	Q9hed0 neurospora
5	104	6.8	1081	2	Q9U631	Q9u631 drosophila
6	103	6.7	255	2	Q6P7N7	Q6p7n7 homo sapien
7	102	6.7	1091	2	Q7YU67	Q7yu67 drosophila
8	102	6.7	1093	2	Q9VTT0	Q9vtt0 drosophila
9	100	6.5	314	2	Q8IHM5	Q8ihm5 plasmodium
10	99	6.5	255	2	Q6UVZ4	Q6uvz4 homo sapien
11	99	6.5	628	2	Q73Q59	Q73q59 treponema d
12	97.5	6.4	456	2	Q9FCY5	Q9fcy5 escherichia
13	97	6.3	709	1	KRP1_SCHPO	Q09175 schizosacch
14	95	6.2	1014	2	Q9V9Q9	Q9v9q9 drosophila
15	95	6.2	1353	2	Q9V9Q7	Q9v9q7 drosophila
16	94	6.2	259	2	Q9D5K1	Q9d5k1 m mus muscu
17	93.5	6.1	262	2	Q6AXW8	Q6axw8 rattus norv
18	93	6.1	794	2	Q60426	Q60426 cricetulus
19	92.5	6.1	797	1	FURI_BOVIN	Q28193 bos taurus
20	92	6.0	1949	2	Q9DF53	Q9df53 brachydanio
21	91	6.0	401	1	CSP_PLACG	P08674 plasmodium
22	91	6.0	1213	2	Q6NZL6	Q6nzl6 mus musculu
23	90.5	5.9	1003	2	Q8TNK1	Q8tnk1 methanosarc
24	90.5	5.9	1643	2	Q91WX9	Q91wx9 rattus norv
25	90.5	5.9	2179	2	Q7TSD2	Q7tsd2 mus musculu
26	90	5.9	985	2	Q7Q8G8	Q7q8g8 anopheles g
27	89.5	5.9	745	2	Q7PIY6	Q7piy6 anopheles g
28	89.5	5.9	800	2	O00835	O00835 dictyosteli
29	89	5.8	226	2	Q26136	Q26136 plasmodium
30	89	5.8	373	2	Q6FU83	Q6fu83 candida gla
31	89	5.8	378	1	CSP_PLACL	P08675 plasmodium
32	88.5	5.8	533	2	Q7XAF7	Q7xaf7 brassica na
33	88.5	5.8	1527	2	Q6PPA4	Q6ppa4 leishmania
34	88	5.8	368	2	Q8MPK1	Q8mpk1 plasmodium
35	88	5.8	456	2	Q659Z2	Q659z2 escherichia
36	88	5.8	1067	2	Q84QV5	Q84qv5 oryza sativ
37	87.5	5.7	793	1	FURI_RAT	P23377 rattus norv
38	87.5	5.7	1098	2	Q6BH26	Q6bh26 debaryomyce
39	87	5.7	1020	2	Q6Q918	Q6q918 uncultured
40	87	5.7	1047	2	Q6JJ55	Q6jj55 ipomoea tri
41	86.5	5.7	271	1	TNR4_RAT	P15725 rattus norv
42	86.5	5.7	302	2	Q9GQL9	Q9gql9 plasmodium
43	86.5	5.7	350	2	Q26117	Q26117 plasmodium
44	86.5	5.7	350	2	Q26118	Q26118 plasmodium
45	86.5	5.7	377	2	Q26116	Q26116 plasmodium

ALIGNMENTS

RESULT 1
Q9HBV2
ID Q9HBV2 PRELIMINARY; PRT; 294 AA.
AC Q9HBV2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Sperm acrosome membrane protein SAMP32 (Sperm acrosome associated
DE 1).
GN Name=SAMP32; Synonyms=SPACA1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21858330; PubMed=11870081;
RA Hao Z., Wolkowicz M.J., Shetty J., Klotz K., Bolling L., Sen B.,
RA Westbrook V.A., Coonrod S., Flickinger C.J., Herr J.C.;
RT "SAMP32, a testis-specific, isoantigenic sperm acrosomal membrane-
RT associated protein.";
RL Biol. Reprod. 66:735-744(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF203447; AAG31422.1; -.
DR EMBL; BC029488; AAH29488.1; -.
DR Genew; HGNC:14967; SPACA1.

```
DR   GO; GO:0008083; F:growth factor activity; IEA.
DR   InterPro; IPR000762; PTN_MK.
SQ   SEQUENCE   294 AA;  32143 MW;  60DB3107E1B03D12 CRC64;

  Query Match             100.0%;  Score 1528;  DB 2;  Length 294;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-122;
  Matches  294;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAE 60

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120

Qy        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLA 180

Qy        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFI 240

Qy        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294


RESULT 2
Q9DA48
ID   Q9DA48      PRELIMINARY;      PRT;   305 AA.
AC   Q9DA48;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE   library, clone:1700021A02 product:weakly similar to SPERM ACROSOME
DE   MEMBRANE PROTEIN SAMP32.
GN   Name=4930540L03Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK006187; BAB24447.1; -.
DR   MGD; MGI:1914902; 4930540L03Rik.
DR   GO; GO:0008083; F:growth factor activity; IEA.
DR   InterPro; IPR000762; PTN_MK.
SQ   SEQUENCE   305 AA;  33342 MW;  9D68F7322E2DB398 CRC64;

  Query Match             73.0%;  Score 1115.5;  DB 2;  Length 305;
  Best Local Similarity   68.6%;  Pred. No. 2.7e-87;
  Matches  210;  Conservative   42;  Mismatches   41;  Indels   13;  Gaps    5;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAA---VQDAGLAHEGEGEEETEN---- 53
              |  || || ||||  ||||:| |||::: :|||::   ||:  :| ||| | |:|:
Db          1 MRARGAGCPAGLL-AVGWLILVGLQASQASNVTSSGGGVQEPAVAREGESESESESEEEA 59

Qy         54 -NDSETAENYAPPET---EDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCV 109
               |: |  |:    |    |:| || :||||||||||||||:|:|||||||||||||||||
Db         60 ENEGEVPESETTAEADAEEEVQNRTIVKEVEFGMCTVTCGVGIREVILTNGCPGGESKCV 119

Qy        110 VRVEECRGPTDCGWGKPISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAI 169
              ||||||||| ||||||||||:|:| ||:|:| || |||||:||||: ||| :|| ||||:
Db        120 VRVEECRGPVDCGWGKPISENLDSARLSCVHISPENRFKYVWKLLKPDQQPVILTNDSAV 179

Qy        170 LEVRKESHPLAFECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVII 229
              ||: :|  |||||||||||||:||::||||||::|||||||| | ||| |:|||||||||
Db        180 LEITREIRPLAFECDTLDNNEMVASVKFTVYTTNELQMRRSSRPDTDAVLVFVLTIGVII 239

Qy        230 CVFIIFLLIFIIINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMP-GEDDA 288
              |:|:||:|||||||||||:|||:| |  |:||| ||:|||||||||| ||::|  |||| 
Db        240 CIFVIFVLIFIIINWAAVKSFWGSKTSATEIQSELSSMRYKDSTSLDQSPTDIPVHEDDA 299

Qy        289 LSEWNE 294
              ||||||
Db        300 LSEWNE 305


RESULT 3
Q80VM2
ID   Q80VM2      PRELIMINARY;      PRT;    90 AA.
AC   Q80VM2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE   Similar to RIKEN cDNA 4930540L03 gene.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Olfactory epithelium;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048388; AAH48388.1; -.
SQ   SEQUENCE   90 AA;  10159 MW;  A049C5010A9E5CC6 CRC64;

  Query Match             20.4%;  Score 311;  DB 2;  Length 90;
  Best Local Similarity   67.4%;  Pred. No. 6.4e-19;
  Matches   58;  Conservative   16;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        191 IVATIKPTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFIIINWAAVKAF 250
              :||::|||||::||||||| | ||| |:||||||||:|:||:||||||||||:|
Db          1 MVASVKPTVYTTNELQMRRSSRPDTDAVLVFVLTIGVIICIFVIFVLIFIIINWAAVKSF 60

Qy        251 WGAKASTPEVQSEQSSVRYKDSTSLD 276
              ||:| |  |:||| ||:| :   ::
Db         61 WGSKTSATEIQSELSSMRCEQDKDIE 86


RESULT 4
Q9HED0
ID   Q9HED0      PRELIMINARY;      PRT;  1331 AA.
AC   Q9HED0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Related to neomycin resistance protein NEO1.
GN   Name=99H12.220;
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL451018; CAC18258.2; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0016887; F:ATPase activity; IEA.
DR   GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0006812; P:cation transport; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR005834; Dehal_like_hydro.
DR   InterPro; IPR006539; Flippase.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
DR   TIGRFAMs; TIGR01494; ATPase_P-type; 3.
DR   PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
SQ   SEQUENCE   1331 AA;  147482 MW;  9FC644FC83820608 CRC64;

  Query Match             7.2%;  Score 110;  DB 2;  Length 1331;
  Best Local Similarity   25.2%;  Pred. No. 2.8;
  Matches   61;  Conservative   38;  Mismatches   93;  Indels   50;  Gaps   14;

Qy         41 LAHEGEGEEETENNDSETAENYA---PPETEDV---SNRNVVKEVEFGMCTVTCGIGVRE 94
              ||||    :||| :    |         : :||   |::|     |
Db        378 LAHESAANKETEEQEVPAKEEMLLLDHVDDDDVGEGSSKNAKTSKESN---------- 425

Qy         95 VILTNGCPGGES-KCVVRVEECRGPTDCGW----GKPISESL---ESVRLACIHTSP--- 143
                 ||  |  | :  :| ::  | ||  |       |:|::|    | |||       |
Db        426 ---NNGAEGSSSGETFIRTDQLDGETD--WKLRLASPLSQNLSTEELVRLRVTAGKPDKR 480

Qy        144 LNRFKYMWKLL--RQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA----TIKP 197
              :| |    :||  |||  |    |     :|:  : ||: : :|   | ::|     |:
Db        481 VNEFVGTLELLPSRQDVMSGAAYNPREGDDVK--AAPLSID-NTAWANTVIASNATTLAV 537

Qy        198 TVYTSSELQMRRSSLPATDAALIF---VLTIGVIICVFIIFLLIFIIINWAAVKAFWGAK 254
              |||  : :   |: |:    :    : ::  |:|    || :|: |   |:: |  ||
Db        538 IVYTGPQTRSALSTSPSRSKTGLLEYEINSLTKILC----FLTLFLSIVLVALEGFSTAK 593

Qy        255 AS 256
              :
Db        594 GN 595


RESULT 5
Q9U631
ID   Q9U631      PRELIMINARY;      PRT;  1081 AA.
AC   Q9U631;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Semaphorin 5C.
GN   Name=Sema-5c;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=20171070; PubMed=10704872; DOI=10.1016/S0925-4773(99)00297-X;
RA   Khare N., Fascetti N., DaRocha S., Chiquet-Ehrismann R.,
RA   Baumgartner S.;
RT   "Expression patterns of two new members of the Semaphorin family in
RT   Drosophila suggest early functions during embryogenesis.";
RL   Mech. Dev. 91:393-397(2000).
DR   EMBL; AF198084; AAF04860.1; -.
DR   HSSP; P07996; 1LSL.
DR   FlyBase; FBgn0028679; Sema-5c.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008199; F:ferric iron binding; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0006725; P:aromatic compound metabolism; IEA.
```

```
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR000183; Decarbxylse2.
DR   InterPro; IPR000627; Dioxygenase.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR001627; Sema.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR008085; TSP_1.
DR   Pfam; PF01437; PSI; 1.
DR   Pfam; PF01403; Sema; 1.
DR   Pfam; PF00090; TSP_1; 6.
DR   PRINTS; PR01705; TSP1REPEAT.
DR   SMART; SM00423; PSI; 1.
DR   SMART; SM00630; Sema; 1.
DR   SMART; SM00209; TSP1; 6.
DR   PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR   PROSITE; PS50092; TSP1; 6.
SQ   SEQUENCE   1081 AA;  120435 MW;  561071C831C431D3 CRC64;

  Query Match             6.8%;  Score 104;  DB 2;  Length 1081;
  Best Local Similarity   22.1%;  Pred. No. 7.2;
  Matches   62;  Conservative   35;  Mismatches  103;  Indels   80;  Gaps   12;

Qy          43 HEGEGEEETENNDSETAENYAPPETEDVSNRNVVKEVEFG---------MCTVTCGIGVR 93
               | | | :      |     :| |  | :|: ::  ||| |         |:||||:|:|
Db         806 HRGRGSQ------SRVCNMHACPAEEQLSSNSLDNEVEHGEWGCWSEWSACSVTCGLGLR 859

Qy          94 EVILTNGCPGG-----------ESKC-VVRVEECRGPTDCGWGKPISESLESVRLACIHT 141
                 |  | |             | || :|  |   | :   | :  | | | : :||
Db         860 R--RTRRCLAGHDRLCQGRALEEQKCEMVPCENFLGWS--AWSEWSSCSSDGIRL----- 910

Qy         142 SPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVATIKFTVYT 201
                  ::   |: |                    | : ||      |  : |
Db         911 ------RHRRCLVEQ--------------------PGSMECRG-------AEFEKTACV 936

Qy         202 SSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLI---FIIINWAAVKAFWGAKASTP 258
                :| :  :::  ||   :|||  :  : |    :       |::    |:       ||
Db         937 PNECEETQTASTATLPIVIFVGLLFTVACCLATYRFTKKRFMLSAEEALNKTTTTTASFD 996

Qy         259 EVQSEQSSVRYKDSTSLDQLPT-----EMPGEDDALSEWN 293
                 :: ||:  ||    || |       || :   |   |
Db         997 TYPNQYSSLPTKD--YYDQRPKRQSSFRMPAKTSNLGNGN 1034


RESULT 6
Q6P7N7
ID   Q6P7N7      PRELIMINARY;      PRT;   255 AA.
AC   Q6P7N7;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Similar to RIKEN cDNA 4930429O20.
GN   Name=LOC388730;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral Nervous System;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral Nervous System;
RA   Strausberg R.;
RL   Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC061592; AAH61592.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   PROSITE; PS50835; IG_LIKE; 1.
SQ   SEQUENCE   255 AA;  28468 MW;  27FDEBE4FA76E021 CRC64;

  Query Match             6.7%;  Score 103;  DB 2;  Length 255;
  Best Local Similarity   24.8%;  Pred. No. 1.5;
  Matches   41;  Conservative   21;  Mismatches   65;  Indels   38;  Gaps    8;

Qy          58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
               | :  | ||    :   |: :    ||||||:| : | :   |  |   ||  |  ||
Db          25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTRRLECL 82

Qy         117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                 ||          ||          |:|: :  |      |:: |:| |     :|  :
Db          83 TNWICGMLHFTILIGK-------EFELSCLSSDILEPGQEAFRFTWRLAR----GVISTD 131

Qy         166 DSAILEVRKESHPLAFE-----------CDT--LDNNEIVATIKF 197
               |      :  || : |:           ||   : |  :|  : |
Db         132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 7
Q7YU67
ID   Q7YU67      PRELIMINARY;      PRT;  1091 AA.
AC   Q7YU67;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   RE68041p.
GN   Name=Sema-5c;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (AUG-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT009966; AAQ22435.1; -.
DR   HSSP; Q92854; 1OLZ.
DR   FlyBase; FBgn0028679; Sema-5c.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008199; F:ferric iron binding; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0006725; P:aromatic compound metabolism; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR000183; Decarbxylse2.
DR   InterPro; IPR000627; Dioxygenase.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR001627; Sema.
```

```
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF01437; PSI; 1.
DR   Pfam; PF01403; Sema; 1.
DR   Pfam; PF00090; TSP_1; 6.
DR   SMART; SM00423; PSI; 1.
DR   SMART; SM00630; Sema; 1.
DR   SMART; SM00209; TSP1; 6.
DR   PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR   PROSITE; PS50092; TSP1; 6.
SQ   SEQUENCE   1091 AA;  121681 MW;  2DFA873668BACF35 CRC64;

  Query Match             6.7%;  Score 102;  DB 2;  Length 1091;
  Best Local Similarity   21.8%;  Pred. No. 11;
  Matches   61;  Conservative   37;  Mismatches  102;  Indels   80;  Gaps   12;

Qy         43 HEGEGEEETENNDSETAENYAPPETEDVSNRNVVKEVEFG---------MCTVTCGIGVR 93
              | | | :      |     :| |  | :|: ::  :|| |          |:||||:|:|
Db        816 HRGRGSQ------SRVCNIHACPAEEQLSSNSLDNELEHGEWGCWSEWSACSVTCGLGLR 869

Qy         94 EVILTNGCPGG-----------ESKC-VVRVEECRGPTDCGWGKPISESLESVRLACIHT 141
                |  |  |           | || :|  |:  | :   | :  | | : :||
Db        870 R--RTRRCLAGHDRLCQGRALEEQKCEMVPCEDFLGWS--AWSEWSSCSSDGIRL----- 920

Qy        142 SPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVATIKFTVYT 201
                ::   |: |                          | : ||         |  : |
Db        921 ------RHRRCLVEQ--------------------PGSMECRG-------AEFEKTACV 946

Qy        202 SSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLI---FIIINWAAVKAFWGAKASTP 258
               :| :  ::: ||   :|||  :  : |    :       |::     |:        ||
Db        947 PNECEETQTASTATLPIVIFVGLLFTVACCLATYRFTKKRFMLSAEEALNKTTTTTASFD 1006

Qy        259 EVQSEQSSVRYKDSTSLDQLPT-----EMPGEDDALSEWN 293
              :: ||:  ||    || |        || :   |   |
Db       1007 TYPNQYSSLPTKD--YYDQRPKRQSSFRMPAKTSNLGNGN 1044


RESULT 8
Q9VTT0
ID   Q9VTT0      PRELIMINARY;      PRT;  1093 AA.
AC   Q9VTT0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CG5661-PA.
GN   Name=Sema-5c; ORFNames=CG5661;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003542; AAF49966.3; -.
DR   HSSP; P07996; 1LSL.
DR   IntAct; Q9VTT0; -.
DR   FlyBase; FBgn0028679; Sema-5c.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008199; F:ferric iron binding; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0006725; P:aromatic compound metabolism; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR000183; Decarbxylse2.
DR   InterPro; IPR000627; Dioxygenase.
DR   InterPro; IPR003659; Plexin-like.
```

```
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR001627; Sema.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR008085; TSP_1.
DR   Pfam; PF01437; PSI; 1.
DR   Pfam; PF01403; Sema; 1.
DR   Pfam; PF00090; TSP_1; 6.
DR   PRINTS; PR01705; TSP1REPEAT.
DR   SMART; SM00423; PSI; 1.
DR   SMART; SM00630; Sema; 1.
DR   SMART; SM00209; TSP1; 6.
DR   PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR   PROSITE; PS50092; TSP1; 6.
SQ   SEQUENCE   1093 AA;  121927 MW;  8A36FE1A22FB1DBF CRC64;

  Query Match             6.7%;  Score 102;  DB 2;  Length 1093;
  Best Local Similarity   21.8%;  Pred. No. 11;
  Matches   61;  Conservative   37;  Mismatches  102;  Indels   80;  Gaps   12;

Qy          43 HEGEGEEETENNDSETAENYAPPETEDVSNRNVVKEVEFG---------MCTVTCGIGVR 93
               | | | :       |     :| |  | :|: ::  |:| |          |:||||:|:|
Db         818 HRGRGSQ------SRVCNIHACPAEEQLSSNSLDNELEHGEWGCWSEWSACSVTCGLGLR 871

Qy          94 EVILTNGCPGG----------ESKC-VVRVEECRGPTDCGWGKPISESLESVRLACIHT 141
                  |  |  |           | || :|  |:  | :   | :  | | : :||
Db         872 R--RTRRCLAGHDRLCQGRALEEQKCEMVPCEDFLGWS--AWSEWSSCSSDGIRL----- 922

Qy         142 SPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAPECDTLDNNEIVATIKFTVYT 201
                  ::   |: |                          | : ||         |  : |
Db         923 ------RHRRCLVEQ--------------------PGSMECRG-------AEFEKTACV 948

Qy         202 SSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLI---FIIINWAAVKAFWGAKASTP 258
                :| :  :::  ||   :||| : :  : |   :      |::     |:      ||
Db         949 PNECEETQTASTATLPIVIFVGLLFTVACCLATYRFTKKRFMLSAEEALNKTTTTTASPD 1008

Qy         259 EVQSEQSSVRYKDSTSLDQLPT-----EMPGEDDALSEWN 293
                  :: ||:  ||    || |       || :   |   |
Db        1009 TYPNQYSSLPTKD--YYDQRPKRQSSFRMPAKTSNLGNGN 1046


RESULT 9
Q8IHM5
ID   Q8IHM5      PRELIMINARY;      PRT;   314 AA.
AC   Q8IHM5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Rifin.
GN   ORFNames=PF11_0515;
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014843; AAN36084.1; -.
DR   InterPro; IPR011038; Calycin.
DR   InterPro; IPR006373; Rifin.
DR   InterPro; IPR002858; Rifin_STEVOR.
DR   Pfam; PF02009; Rifin_STEVOR; 1.
DR   TIGRFAMs; TIGR01477; RIFIN; 1.

SQ   SEQUENCE   314 AA;  35249 MW;  F2F02922448AD040 CRC64;

  Query Match             6.5%;  Score 100;  DB 2;  Length 314;
  Best Local Similarity   22.4%;  Pred. No. 3.6;
  Matches   56;  Conservative   36;  Mismatches   92;  Indels   66;  Gaps    9;

Qy          52 ENNDSETAENYAPPETEDVSNRNVVKE--------------VE--------------FG 82
               || | :|:: :   |   : ||   ||              :|              ||
Db          64 ENFDRQTSQRFEEYEERMIKNRKRCKEQCDKDIKKIIVKDKIEKSFAEKLEKGCLRCGFG 123

Qy          83 MCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTDCGWGKPIS--------ESLESV 134
               :  |   :|:   |: |     | |     |   :  |: |  | |          | |
Db         124 LGGVAASVGIIGPIIVN-----ELKKTALVAAAQTGTEAGIDKAIEVVISKYGVNKLYGV 178

Qy         135 RLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAPECDTLDNNEIV-- 192
                 |    ||  | || :   ::  |    :  ||  |   :  || |  |:|::  :
Db         179 ALEKSITS--NNFKNVMFYIQAIQHRYNTMVCSA--EPIDDIGPLCFLKDSLNDGVLFTK 234

Qy         193 ------------ATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGV--IICVFIIFLLI 238
                         || | |: | :|:    |:  ||    | : :   |  |: : :: ::|
Db         235 SISASAQKVVADATEKATLVTKAEV----SAAEATSVNLYYAIAYSVIAILVIVLVMVII 290

Qy         239 FIIINWAAVK 248
               |:|: :   |
Db         291 FLILRYRRKK 300


RESULT 10
Q6UVZ4
ID   Q6UVZ4      PRELIMINARY;      PRT;   255 AA.
AC   Q6UVZ4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   KVLA2788.
GN   ORFNames=UNQ2788;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22887296; PubMed=12975309; DOI=10.1101/gr.1293003;
RA   Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,
RA   Chen J., Chow B., Chui C., Crowley C., Currell B., Deuel B., Dowd P.,
RA   Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.E., Heldens S.,
RA   Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,
RA   Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,
RA   Seshagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,
RA   Vandlen R., Watanabe C., Wieand D., Woods K., Xie M.H., Yansura D.,
RA   Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,
RA   Godowski P.;
RT   "The secreted protein discovery initiative (SPDI), a large-scale
RT   effort to identify novel human secreted and transmembrane proteins: a
RT   bioinformatics assessment.";
RL   Genome Res. 13:2265-2270(2003).
DR   EMBL; AY359081; AAQ89440.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   PROSITE; PS50835; IG_LIKE; 1.
SQ   SEQUENCE   255 AA;  28440 MW;  E7FEEAAF968A1E95 CRC64;

  Query Match             6.5%;  Score 99;  DB 2;  Length 255;
  Best Local Similarity   24.2%;  Pred. No. 3.4;
  Matches   40;  Conservative   22;  Mismatches   65;  Indels   38;  Gaps    8;

Qy          58 TAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVR-EVILTNGCPGGESKCVVRVEECR 116
               | :  | ||    :   |:  :    ||||||:| : | :   |  |   ||  :  ||
Db          25 TPKTLAIPEKLQEAVGKVI--INATTCTVTCGLGYKEETVCEVGPDGVRRKCQTQRLECL 82

Qy         117 GPTDCGW-------GKPISESLESVRLACIHTSPL----NRFKYMWKLLRQDQQSIILVN 165
                  ||          ||          |:|: :  |       |:: |:| |     :|  :
Db          83 TNWICGMLHFTILIGK-------EFELSCLSSDILEFGQEAFRFTWRLAR----GVISTD 131
```

```
Qy        166 DSAILEVRKESHPLAFE----------CDT--LDNNEIVATIKF 197
              |       :  || : |:          ||   : |  :|  : |
Db        132 DEVFKPFQANSHFVKFKYAQEYDSGTYRCDVQLVKNLRLVKRLYF 176


RESULT 11
Q73Q59
ID   Q73Q59      PRELIMINARY;      PRT;   628 AA.
AC   Q73Q59;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=TDE0585;
OS   Treponema denticola.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=158;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35405 / DSM 14222;
RX   PubMed=15064399; DOI=10.1073/pnas.0307639101;
RA   Seshadri R., Myers G.S.A., Tettelin H., Eisen J.A., Heidelberg J.F.,
RA   Dodson R.J., Davidsen T.M., DeBoy R.T., Fouts D.E., Haft D.H.,
RA   Selengut J., Ren Q., Brinkac L.M., Madupu R., Kolonay J.F.,
RA   Durkin S.A., Daugherty S.C., Shetty J., Shvartsbeyn A.,
RA   Gebregeorgis B., Geer K., Tsegaye G., Malek J.A., Ayodeji B.,
RA   Shatsman S., McLeod M.P., Smajs D., Howell J.K., Pal S., Amin A.,
RA   Vashisth P., McNeill T.Z., Xiang Q., Sodergren E., Baca E.,
RA   Weinstock G.M., Norris S.J., Fraser C.M., Paulsen I.T.;
RT   "Comparison of the genome of the oral pathogen Treponema denticola
RT   with other spirochete genomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:5646-5651(2004).
DR   EMBL; AE017248; AAS11080.1; -.
DR   TIGR; TDE0585; -.
KW   Complete proteome.
SQ   SEQUENCE   628 AA;  72538 MW;  8091B1A1BB9AAB42 CRC64;

  Query Match             6.5%;  Score 99;  DB 2;  Length 628;
  Best Local Similarity   21.1%;  Pred. No. 10;
  Matches   64;  Conservative   35;  Mismatches  106;  Indels   98;  Gaps   10;

Qy         32 VTAAVQDAGLAHEGEGEEETENN----DSETAENYAPPETEDVSNRNVVKEVEFGMCTVT 87
              :|       |:  :|  |||  |     : ::| |  |:                : | |
Db        307 ITQFTDSRGVPLLGQYSEETLQNFMKEKASSSEKYIIPD---------------ILTFT 350

Qy         88 CGIGVREVILTNGCPGGESKCVVRVEECRGPTDCGWGKPISESLESVRLACIHTSPLNRF 147
                 | :||        | |  :          |:|: : ||  ||
Db        351 NSSNDRFYLLT-------EKVVPLLISL-----------INEARKPVRELCI-------- 384

Qy        148 KYMWKLLRQDQQSIILVNDSAILEVRKE----------------------SHPLAFECDT 185
              |  ::|  .:|   : ||:|  |: ||                        : |   |
Db        385 KKWHEMLMNFEQDDSMKNDTAFNELLKEITAHSAPNLYGLLNAPFILSIIADPRLNEIQA 444

Qy        186 LDNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLIFIIINWA 245
              :: | |    |   |:   :  |  |  | |  || :|                 :
Db        445 MEINRIFPAGKPAAYSEILMLNRHEILSDTKILLPFWYTIPII----------------S 488

Qy        246 AVKAFWGAKASTPE-VQSEQSSVRYK----------DSTSLDQLPTEMPGE---DDALSE 291
               : ||:  |  ||: || |:    ||          :  | : :|  |  |   :  | |
Db        489 TIIAPFKRKKKTPQPVQPEKKETPYKKPKLKLKDAAEKISAEFIPQGMTMEQALEKNLDE 548

Qy        292 WNE 294
              ||:
Db        549 WNQ 551


RESULT 12
Q9FCY5
ID   Q9FCY5      PRELIMINARY;      PRT;   456 AA.
AC   Q9FCY5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Invasion protein IbeA.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96029746; PubMed=7591087;
RA   Huang S.H., Wass C., Fu Q., Prasadarao N.V., Stins M., Kim K.S.;
RT   "Escherichia coli invasion of brain microvascular endothelial cells in
RT   vitro and in vivo: molecular cloning and characterization of invasion
RT   gene ibe10.";
RL   Infect. Immun. 63:4470-4475(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20465456; PubMed=11008113; DOI=10.1016/S1286-4579(00)01277-6;
RA   Huang S.H., Stins M.F., Kim K.S.;
RT   "Bacterial penetration across the blood-brain barrier during the
RT   development of neonatal meningitis.";
RL   Microbes Infect. 2:1237-1244(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21136444; PubMed=11237832;
RA   Huang S.H., Wan Z.S., Chen Y.H., Jong A.Y., Kim K.S.;
RT   "Further characterization of Escherichia coli brain microvascular
RT   endothelial cell invasion gene ibeA by deletion, complementation, and
RT   protein expression.";
RL   J. Infect. Dis. 183:1071-1078(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21652705; PubMed=11793250;
RA   Huang S.H., Chen Y.H., Kong G., Chen S.H.M., Besemer J.,
RA   Borodovsky M., Jong A.;
RT   "A novel genetic island of meningitic Escherichia coli K1 containing
RT   the ibeA invasion gene (GimA): functional annotation and carbon-
RT   source-regulated invasion of human brain microvascular endothelial
RT   cells.";
RL   Funct. Integr. Genomics 1:312-322(2001).
DR   EMBL; AF289032; AAF98391.2; -.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR000205; NAD_BS.
DR   InterPro; IPR000103; Pyridine_redox_2.
DR   PRINTS; PR00469; PNDRDTASEII.
SQ   SEQUENCE   456 AA;  49772 MW;  00FD72D7E9B0E010 CRC64;

  Query Match             6.4%;  Score 97.5;  DB 2;  Length 456;
  Best Local Similarity   18.9%;  Pred. No. 9.2;
  Matches   53;  Conservative   39;  Mismatches   96;  Indels   93;  Gaps   10;

Qy          7 GCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAENYAPPE 66
              ||  |:: |      ||::|         |:  ||| | |          |||::
Db         49 GCFGGMMTT------AGVESIAWWRHENTVESGGLAREIE----------ETAKSMGASS 92

Qy         67 TEDVSNRNVVKEVEFGMC--TVTCGIGVREV----------------ILTNGCPGGESK 107
                | ||   : |  | :    :    |||  |                  ::|    | ::
Db         93 PEPQSNSQAINEERFKLVADAMLEQAGVRRVLHITAVDVIKQGNNLLGVITESKSGRQAI 152

Qy        108 CVVRVEECRGPTDCGW--GKPISESLESVRLAC-------------------------- 138
                : :| |  |  |  | |  :  |   | |
Db        153 LANVIIDCTGDADIAWFAGAPFIKR-EREELMCMTTVFSCANINKNAFMQNINSTEPKYG 211

Qy        139 -------------IHTSPLNRFK-YMWKLLRQDQQSIILVND-------SAILEVRKES 176
                          :| |  : |  |: |:||::: : |:  |         | : |    :
Db        212 DWGADEENKNWSYDVHESCRDMFSPYLGKVLRKERSAGIIPKDVTLGGSWSTVTEYGDAN 271

Qy        177 HPLAFECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDA 217
              :         :|  ::          | :|:: |: :: | :|
Db        272 YLNVVSIPAVDCTDVFD------LTRAEIEGRKQAMQAIEA 306


RESULT 13
KRP1_SCHPO
```

```
ID   KRP1_SCHPO     STANDARD;      PRT;   709 AA.
AC   Q09175;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Dibasic processing endoprotease precursor (EC 3.4.21.-) (KEX2-related
DE   protease).
GN   Name=krp1; Synonyms=krp; ORFNames=SPAC22E12.09c;
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EG545;
RX   MEDLINE=95112801; PubMed=7813430;
RA   Davey J., Davis K., Imai Y., Yamamoto M., Matthews G.;
RT   "Isolation and characterization of krp, a dibasic endopeptidase
RT   required for cell viability in the fission yeast Schizosaccharomyces
RT   pombe.";
RL   EMBO J. 13:5910-5921(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Membrane-bound, subtilisin-like serine protease that
CC       processes the P-factor precursor and other precursor proteins.
CC       Essential for cell viability.
CC   -!- CATALYTIC ACTIVITY: Cleaves substrate on the C-terminal side of
CC       dibasic residues.
CC   -!- COFACTOR: Calcium.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Late Golgi
CC       compartment (By similarity).
CC   -!- PTM: N-glycosylated.
CC   -!- SIMILARITY: Belongs to the peptidase S8 family. Furin subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82435; CAA57818.1; -.
DR   EMBL; Z70043; CAA93896.1; -.
DR   PIR; S51793; S51793.
DR   HSSP; P13134; 1OT5.
DR   MEROPS; S08.070; -.
DR   GeneDB_SPombe; SPAC22E12.09c; -.
DR   InterPro; IPR000209; Pept_S8_S53.
DR   InterPro; IPR002884; PrprotnconvertsP.
DR   Pfam; PF01483; P_proprotein; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   ProDom; PD000717; PrprotnconvertsP; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Calcium; Glycoprotein; Hydrolase; Serine protease; Signal;
KW   Transmembrane; Zymogen.
FT   SIGNAL       1     22       Potential.
FT   PROPEP      23     82       Potential.
FT   PROPEP      83    102       Potential.
FT   CHAIN      103    709       Dibasic processing endoprotease.
FT   DOMAIN     103    668       Lumenal (Potential).
FT   TRANSMEM   669    693       Potential.
FT   DOMAIN     694    709       Cytoplasmic (Potential).
FT   DOMAIN     612    647       Ser/Thr-rich.
FT   ACT_SITE   162    162       Charge relay system (By similarity).
FT   ACT_SITE   200    200       Charge relay system (By similarity).
FT   ACT_SITE   371    371       Charge relay system (By similarity).
FT   DISULFID   216    363       By similarity.
FT   DISULFID   308    338       By similarity.
FT   CARBOHYD   155    155       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD   463    463       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD   471    471       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD   620    620       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   709 AA;  78126 MW;  414FE9B89CBE0840 CRC64;

  Query Match             6.3%;  Score 97;  DB 1;  Length 709;
  Best Local Similarity   25.8%;  Pred. No. 17;
  Matches   40;  Conservative   17;  Mismatches   52;  Indels   46;  Gaps    7;

Qy        150 MWKLLRQDQQ----------SIILVNDSAILEVRKESHPLAFECDTLD---------- 187
              :||||  |:            : |  :|   |    : || : |||:
Db        558 VWKLLVNDRSGGKHEGTFENWQLALWGES---ENPSNTAPLPY--DTLELPKEMVLGIYS 612

Qy        188 --NNEIV--------ATIKFTVYTSSELQMRRSS--------LPATDAAL-IFVLTIGVI 228
              |:::           :  || || |      |:         || |   |     |
Db        613 EPNSDLTNSSTLLSPTSTSFTSYTVSATATPTSTSHIPIPTVLPPTQPVLEPSYREIVAF 672

Qy        229 ICVFIIFLLIFIIINWAAVKAFWGAKASTPEVQSE 263
              |  |::|  ||: : |  : ||| |||  |  | |
Db        673 ITFFLLFAFIFVAVIWTWISAFWKAKAPPPLSQQE 707


RESULT 14
Q9V9Q9
ID   Q9V9Q9      PRELIMINARY;      PRT;  1014 AA.
AC   Q9V9Q9; Q95R33;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   CG31619-PB (GH19218p).
GN   Name=CG2131; ORFNames=CG31619;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
```

```
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003781; AAF57227.2; -.
DR   EMBL; AY061825; AAL27636.1; -.
DR   FlyBase; FBgn0051619; CG31619.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR008085; TSP_1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00090; TSP_1; 7.
DR   PRINTS; PR01705; TSP1REPEAT.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00209; TSP1; 8.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS50092; TSP1; 6.
SQ   SEQUENCE   1014 AA;  112240 MW;  030A1645935D5360 CRC64;

  Query Match             6.2%;  Score 95;  DB 2;  Length 1014;
  Best Local Similarity   21.9%;  Pred. No. 39;
  Matches   61;  Conservative   33;  Mismatches   88;  Indels   96;  Gaps   14;

Qy          8 CSAGL-----LMTVGWLLLAGLQSARGTNVTAAVQ----DAGLAHEGEGEEETENNDSET 58
              || |:     |  ||  :: |  |:|  |  :  |    ||   :| | :|  : :| |:
Db        419 CSTGIHCGGSLNKVGGTIIVG--SSRSLNERSERQLDSSDADEDNEDENDEGDDVDDLES 476

Qy         59 AEN--------------YA-------------PPETEDVSNRN---------------- 74
               ::              ||                 | |  : | |
Db        477 GQDTDDGEGLSYADQPLLYAHRTQSRLNQEAPDEPRTMHLMNGNSNNNFNRGEDESEGPS 536

Qy         75 ----VVKEVEFGMCTVTCGIGVREVILTNGCP--GGESKCVVRVEE--CRGPTDCGWGKP 126
                  :|: |:  |:|||| |:|    |  |       |:|  |  | :  | |       ||
Db        537 LDPTYIKDNEWSPCSVTCGEGIRR--RTYNCKIFLEYSRTVATVNDSLCEGK------KP 588

Qy        127 ISESLESVRLACIHTS-------PLNRFK---------YMWKLLRQDQQSIILVNDSAIL 170
                |    |  |:  |        | :  |         |:|:     :|     : | :
Db        589 HDEVERCVEDPCMLPSHGFDDQFPRDSIKVGVSEPGKTYVWR-----EQGYTSCSASCLG 643

Qy        171 EVRKESHPLAFECDTLDNNEIVATIKFTVYTSSELQMR 208
              | :    |   |  ||  :|:    :  |  | ::|
Db        644 GVEE----LIINCVREDNGRVVSPFLCSPETKPEARVR 677


RESULT 15
Q9V9Q7
ID   Q9V9Q7      PRELIMINARY;      PRT;  1353 AA.
AC   Q9V9Q7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CG31619-PA.
GN   ORFNames=CG31619;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
```

```
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003781; AAF57229.2; -.
DR   FlyBase; FBgn0051619; CG31619.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR010909; PLAC.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00090; TSP_1; 8.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00209; TSP1; 11.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS50900; PLAC; 1.
DR   PROSITE; PS50092; TSP1; 8.
SQ   SEQUENCE   1353 AA;  150210 MW;  F83CDB090964272F CRC64;

  Query Match             6.2%;  Score 95;  DB 2;  Length 1353;
  Best Local Similarity   21.9%;  Pred. No. 55;
  Matches   61;  Conservative   33;  Mismatches   88;  Indels   96;  Gaps   14;

Qy          8 CSAGL-----LMTVGWLLLAGLQSARGTNVTAAVQ----DAGLAHEGEGEEETENNDSET 58
              || |:      |  ||  :: |  |:|  |  :  |    ||   :| | :|  : :| |:
Db        419 CSTGIHCGGSLNKVGGTIIVG--SSRSLNERSERQLDSSDADEDNEDENDEGDDVDDLES 476

Qy         59 AEN--------------YA-------------PPETEDVSNRN--------------- 74
               ::              ||                | |  : | |
Db        477 GQDTDDGEGLSYADQPLLYAHRTQSRLNQEAPDEPRTMHLMNGNSNNNFNRGEDESEGPS 536

Qy         75 ----VVKEVEPGMCTVTCGIGVREVILTNGCP--GGESKCVVRVEE--CRGPTDCGWGKP 126
                  :|: |: |:|||| |:|    | |       |: |  | :  | |       ||
Db        537 LDPTYIKDNEWSPCSVTCGEGIRR--RTYNCKIPLEYSRTVATVNDSLCEGK------KP 588

Qy        127 ISESLESVRLACIHTS-------PLNRFK---------YMWKLLRQDQQSIILVNDSAIL 170
              |    |  |: |        | :  |           |:|:     :|    : | :
Db        589 HDEVERCVEDPCMLPSHGFDDQFPRDSIKVGVSEPGKTYVWR-----EQGYTSCSASCLG 643

Qy        171 EVRKESHPLAPECDTLDNNEIVATIKFTVYTSSELQMR 208
              | :   |  |  || :|:   : |  | ::|
Db        644 GVEE----LIINCVREDNGRVVSPFLCSPETKPEARVR 677


Search completed: April 15, 2005, 14:15:15
Job time : 192 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 15, 2005, 14:08:19 ; Search time 22 Seconds
(without alignments)
997.583 Million cell updates/sec

Title: US-10-809-655-9
Perfect score: 1528
Sequence: 1 MSPRGTGCSAGLLMTVGWLL.....LDQLPTEMPGEDDALSEWNE 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	106	6.9	479	4	US-09-270-767-46823	Sequence 46823, A
2	89.5	5.9	2181	4	US-09-949-016-5981	Sequence 5981, Ap
3	89.5	5.9	2188	4	US-09-949-016-8295	Sequence 8295, Ap
4	85	5.6	467	3	US-09-046-736-2	Sequence 2, Appli
5	85	5.6	596	4	US-09-949-016-7821	Sequence 7821, Ap
6	83.5	5.5	301	5	PCT-US95-13975-72	Sequence 72, Appl
7	83.5	5.5	618	4	US-08-311-731A-158	Sequence 158, App
8	83	5.4	519	4	US-09-248-796A-14534	Sequence 14534, A
9	82.5	5.4	366	3	US-09-134-001C-4799	Sequence 4799, Ap
10	80	5.2	566	4	US-09-513-783A-142	Sequence 142, App
11	80	5.2	566	4	US-09-430-656-142	Sequence 142, App
12	80	5.2	794	1	US-07-885-972A-2	Sequence 2, Appli
13	80	5.2	794	1	US-07-885-972A-4	Sequence 4, Appli
14	80	5.2	794	2	US-08-865-203-2	Sequence 2, Appli
15	80	5.2	794	2	US-08-745-880-2	Sequence 2, Appli
16	80	5.2	794	2	US-08-745-880-4	Sequence 4, Appli
17	80	5.2	794	2	US-08-480-382-2	Sequence 2, Appli
18	80	5.2	794	2	US-08-480-382-4	Sequence 4, Appli
19	80	5.2	794	2	US-07-849-420-2	Sequence 2, Appli
20	80	5.2	794	3	US-09-253-854-2	Sequence 2, Appli
21	80	5.2	794	3	US-08-955-424-2	Sequence 2, Appli
22	80	5.2	794	4	US-09-592-480-1	Sequence 1, Appli
23	80	5.2	794	4	US-10-133-910-2	Sequence 2, Appli
24	80	5.2	794	4	US-09-949-016-6239	Sequence 6239, Ap
25	80	5.2	812	4	US-09-513-783A-4	Sequence 4, Appli
26	80	5.2	812	4	US-09-513-783A-6	Sequence 6, Appli
27	80	5.2	812	4	US-09-430-656-4	Sequence 4, Appli
28	80	5.2	812	4	US-09-430-656-6	Sequence 6, Appli
29	80	5.2	820	4	US-09-949-016-9924	Sequence 9924, Ap
30	80	5.2	1056	4	US-09-513-783A-32	Sequence 32, Appl
31	80	5.2	1056	4	US-09-430-656-32	Sequence 32, Appl
32	80	5.2	1125	4	US-09-513-783A-152	Sequence 152, App
33	80	5.2	1125	4	US-09-430-656-152	Sequence 152, App
34	80	5.2	1610	4	US-09-513-783A-22	Sequence 22, Appl
35	80	5.2	1610	4	US-09-430-656-22	Sequence 22, Appl
36	80	5.2	1858	4	US-09-902-540-12643	Sequence 12643, A
37	79	5.2	582	3	US-09-419-459-2	Sequence 2, Appli
38	78.5	5.1	546	3	US-08-935-855-20	Sequence 20, Appl
39	78.5	5.1	546	4	US-09-538-092-827	Sequence 827, App
40	78.5	5.1	546	4	US-09-949-016-6265	Sequence 6265, Ap
41	78.5	5.1	553	4	US-09-949-016-7397	Sequence 7397, Ap
42	78	5.1	546	4	US-09-540-236-2574	Sequence 2574, Ap
43	78	5.1	935	1	US-07-707-367-2	Sequence 2, Appli
44	78	5.1	1096	4	US-09-540-236-2620	Sequence 2620, Ap
45	77.5	5.1	381	4	US-09-673-395A-363	Sequence 363, App

ALIGNMENTS

RESULT 1
US-09-270-767-46823
; Sequence 46823, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46823
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46823

Query Match 6.9%; Score 106; DB 4; Length 479;
Best Local Similarity 22.1%; Pred. No. 0.0085;
Matches 62; Conservative 36; Mismatches 102; Indels 80; Gaps 12;

```
Qy          43 HEGEGEEETENNDSETAENYAPPETEDVSNRNVVKEVEFG---------MCTVTCGIGVR 93
               | | | :       |     :| |  | :|: ::  ||| |          |:||||:|:|
Db         204 HRGRGSQ------SRVCNMHACPAEEQLSSNSLDNEVEHGEWGCWSEWSACSVTCGLGLR 257

Qy          94 EVILTNGCPGG-----------ESKC-VVRVEECRGPTDCGWGKPISESLESVRLACIHT 141
                 |  |  |             | || :|  |:  | :   | :  | | : :||
Db         258 R--RTRRCLAGHDRLCQGRALEEQKCEMVPCEDFLGWS--AWSEWSSCSSDGIRL----- 308

Qy         142 SPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVATIKFTVYT 201
                     ::   |: |                        | : ||         |  : |
Db         309 ------RHRRCLVEQ--------------------PGSMECRG-------AEFEKTACV 334

Qy         202 SSELQMRRSSLPATDAALIFVLTIGVIICVFIIFLLI---FIIINWAAVKAFWGAKASTP 258
                :| :  ::: ||   :||| :  : |    :       |::    |:       ||
Db         335 PNECEETQTASTATLPIVIFVGLLFTVACCLATYRFTKKRFMLSAEEALNKTTTTTASFD 394

Qy         259 EVQSEQSSVRYKDSTSLDQLPT-----EMPGEDDALSEWN 293
                :: ||:  ||    || |       || :   |   |
Db         395 TYPNQYSSLPTKD--YYDQRPKRQSSFRMPAKTSNLGNGN 432
```

RESULT 2
US-09-949-016-5981
; Sequence 5981, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5981
;   LENGTH: 2181
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-5981

  Query Match             5.9%;  Score 89.5;  DB 4;  Length 2181;
  Best Local Similarity   21.2%;  Pred. No. 6.1;
  Matches   71;  Conservative   46;  Mismatches   97;  Indels  121;  Gaps   18;

Qy         32 VTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIG 91
              :| | :|    :| || || : | |          ||| |:  ||  | |
Db        442 ITQA-EDIDPENEEEGGEEGKRNTSMPTS-----ETESVNTENVSGEGE---------- 484

Qy         92 VREVILTNGCPG---------GESKCVVRVEECRGPTDC-GWGKPISESLESVR------ 135
                 || |          | :|         ||: |  ||: ||:|  | |
Db        485 ------NRGCCGSLWCWWRRRGAAKA--------GPSGCRRWGQAISKSKLSRRWRRWNR 530

Qy        136 ---------------------LACIHTSPLNRFKYM---WKLLRQD-QQSIILVNDSAIL 170
                                  |  ::|  ::   |    |    ||    ::|   :  :
Db        531 FNRRRCRAAVKSVTFYWLVIVLVFLNTLTISSEHYNQPDWLTQIQDIANKVLLALFTCEM 590

Qy        171 EVRKESHPL---------AFEC---------DTLDNNEIVATIKFTVYTSSEL----QMR 208
              |:  |  |          |:|          |   ||:: :  :|:    |    ::
Db        591 LVKMYSLGLQAYFVSLFNRFDCFVVCGGITETILVELEIMSPLGISVFRCVRLLRIFKVT 650

Qy        209 R--SSLPATDAALI-FVLTIGVIICVFIIFLLIFIIINWAAVKAFWGAKASTPEVQSEQS 265
              |  :||    |:|:  : :|  :: :  :|::|| ::       :| | :  | |:::|
Db        651 RHWTSLSNLVASLLNSMKSIASLLLLLFLFIIIFSLLG----MQLFGGKFNFDETQTKRS 706

Qy        266 SVRYKDSTSLDQLPTE-------MPGEDDALSEWN 293
                 : |  |         : |||     ||
Db        707 --------TFDNFPQALLTVFQILTGED-----WN 728


RESULT 3
US-09-949-016-8295
; Sequence 8295, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8295
;   LENGTH: 2188
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-8295

  Query Match             5.9%;  Score 89.5;  DB 4;  Length 2188;
  Best Local Similarity   21.2%;  Pred. No. 6.1;
  Matches   71;  Conservative   46;  Mismatches   97;  Indels  121;  Gaps   18;

Qy         32 VTAAVQDAGLAHEGEGEEETENNDSETAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIG 91
              :| | :|    :| || || : | |          ||| |:  ||  | |
Db        449 ITQA-EDIDPENEEEGGEEGKRNTSMPTS-----ETESVNTENVSGEGE---------- 491

Qy         92 VREVILTNGCPG---------GESKCVVRVEECRGPTDC-GWGKPISESLESVR------ 135
                 || |          | :|         ||: |  ||: ||:|  | |
Db        492 ------NRGCCGSLWCWWRRRGAAKA--------GPSGCRRWGQAISKSKLSRRWRRWNR 537

Qy        136 ---------------------LACIHTSPLNRFKYM---WKLLRQD-QQSIILVNDSAIL 170
                                  |  ::|  ::   |    |    ||    ::|   :  :
Db        538 FNRRRCRAAVKSVTFYWLVIVLVFLNTLTISSEHYNQPDWLTQIQDIANKVLLALFTCEM 597

Qy        171 EVRKESHPL---------AFEC---------DTLDNNEIVATIKFTVYTSSEL----QMR 208
              |:  |  |          |:|          |   ||:: :  :|:    |    ::
Db        598 LVKMYSLGLQAYFVSLFNRFDCFVVCGGITETILVELEIMSPLGISVFRCVRLLRIFKVT 657

Qy        209 R--SSLPATDAALI-FVLTIGVIICVFIIFLLIFIIINWAAVKAFWGAKASTPEVQSEQS 265
              |  :||    |:|:  : :|  :: :  :|::|| ::       :| | :  | |:::|
Db        658 RHWTSLSNLVASLLNSMKSIASLLLLLFLFIIIFSLLG----MQLFGGKFNFDETQTKRS 713

Qy        266 SVRYKDSTSLDQLPTE-------MPGEDDALSEWN 293
                 : |  |         : |||     ||
Db        714 --------TFDNFPQALLTVFQILTGED-----WN 735


RESULT 4
US-09-046-736-2
; Sequence 2, Application US/09046736
; Patent No. 6090582
;  GENERAL INFORMATION:
;    APPLICANT:  KIKLY, KRISTINE
;    APPLICANT:  ERICKSON-MILLER, CONNIE
;    TITLE OF INVENTION:  Sialoadhesin Family Member-3
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/046,736
;      FILING DATE:  24-MAR-1998
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/041,885
;      FILING DATE:  02-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-50019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  467 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-046-736-2
```

```
  Query Match             5.6%;  Score 85;  DB 3;  Length 467;
  Best Local Similarity   19.8%;  Pred. No. 1.8;
  Matches   59;  Conservative   39;  Mismatches  116;  Indels   84;  Gaps   11;

Qy          6 TGCSAGLLMTVGWLLLAG---LQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAENY 62
              :||   |  :| |    |   : |  ||:|:
Db        160 SGCFQNLTCSVPWACEQGTPPMISWMGTSVSP---------------------------- 191

Qy         63 APPETEDVSNRNVVKEVEFGMCTVTC-----GIGV---REVILTNGCPGGESKCVVRVEE 114
                | |    |   :: : :    ::||      | ||   | : |    |       | | :
Db        192 PHPSTTRSSVLTLIPQPQHHGTSLTCQVTLPGAGVTTNRTIQLNVSYP--PQNLTVTVFQ 249

Qy        115 CRGPTDCGWGKPISESL---ESVRLAC-IHTSPLNRFKYMWKLLRQDQQSIILVNDSAIL 170
                |       |   | |:   :|:|| | : ::| |  : :|:               | |
Db        250 GEGTASTALGNSSSLSVLEGQSLRLVCAVDSNPPARLSWTWR-------------SLTL 295

Qy        171 EVRKESHPLAFECDTLDNNEIVATIKFTVYTSSELQMRRSSL-------------PATDA 217
                : |:||  |       :|     :||     : |  :  ||             | :
Db        296 YPSQPSNPLVLELQVHLGDE----GEFTCRAQNSLGSQHVSLNLSLQQEYTGKMRPVSGV 351

Qy        218 ALIFVLTIGVIICVPIIFLLIFIIINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSL 275
                |  |    |    ||: | :|||::           : :    ::    |  ||: ::
Db        352 LLGAVGGAGATALVFLSFCVIFIVVR--------SCRKKSARPAADVGDVGMKDANTI 401


RESULT 5
US-09-949-016-7821
; Sequence 7821, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7821
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-7821

  Query Match             5.6%;  Score 85;  DB 4;  Length 596;
  Best Local Similarity   25.7%;  Pred. No. 2.6;
  Matches   43;  Conservative   23;  Mismatches   55;  Indels   46;  Gaps    9;

Qy         46 EGEEETENND-SETAENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGG 104
              ||  || . ::  |:| |    | : |  ::: || :          |    ||  |      |
Db        229 EGPAETVGSEASDTVEGVLPSKQEPENDQGVVSQ---------AGKDDRESKLE-----G 274

Qy        105 ESKCVVRVEECRGPTDCGWGKPISESLESVRLACIHTSPLNRFKYM--------W----- 151
              ::   :: ||   | | |  ||  |  :   | ||   | :|  ::|      |
Db        275 KAAQDIKDEE---PGDLGRPKPECEGYDPNALYCICRQPHNN-RFMICCDRCEEWFHGDC 330

Qy        152 --------KLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNE 190
                     :|| :: :  |  |   ||:|: |:|      :| | |
Db        331 VGISEARGRLLERNGEDYICPN-CTILQVQDETH-----SETADQQE 371


RESULT 6
PCT-US95-13975-72
; Sequence 72, Application PC/TUS9513975
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  McDonell, Michael W.
;    TITLE OF INVENTION:  Recombinant Feline Herpes virus
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM 330 466 DX2
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/13975
;      FILING DATE:  26-OCT-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/329,883
;      FILING DATE:  26-OCT-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  39118-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)278-0400
;      TELEFAX:  (212)391-0525
;  INFORMATION FOR SEQ ID NO:  72:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  301 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-13975-72

  Query Match             5.5%;  Score 83.5;  DB 5;  Length 301;
  Best Local Similarity   19.8%;  Pred. No. 1.3;
  Matches   59;  Conservative   43;  Mismatches   99;  Indels   97;  Gaps   12;

Qy          4 RGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAV---QDAGLAHEGEGEEETENNDSETAE 60
              || ||    |:  | |:           |::: :       || | : |::    |
Db         81 RGRGCFPELVEPVLWV-----------NISSDLIGGPPFDYNHEDEADIESDELPEEDIH 129

Qy         61 NYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVILTNGCPGGESKCVVRVEECRGPTD 120
              ||:                                                    :||
Db        130 NYS-------------------------------------------------DCRATNM 139

Qy        121 CGWGKPISESLESVRLAC------IHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRK 174
                 :|:|: | | |         | | |       |: |:|::    | |  | || |
Db        140 FVPREPLSQVLGSQSLLVGSLGFQIITQP-------WQ-LKQNESYDGLRN--ASLEPRH 189

Qy        175 ESHPLAFECDTLDNNEIVATIKFTVYTSSELQMRRSSL-------PATDAALIF--VLTI 225
                  : | ||  |:: :|  |   :   :   |         | |:  |:   :: |
Db        190 LDS--SNDRDLLDETEMIGSIITTPPPTHPKGVNGGFLQDLPIIEPTTEPCLVHTKIIGI 247

Qy        226 GVIICVPIIFLLIFIIINWAAVKAFWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMP 283
              | :: ||::|:|| : :    :::  |         ::    ||:  : |  || | |
Db        248 GTVVVVFLLFILISLCVYTCVLRSRIGMVD-----RAYVKQVRFNSNPSYQQL-TRYP 299


RESULT 7
US-08-311-731A-158
; Sequence 158, Application US/08311731A
; Patent No. 6583266
;  GENERAL INFORMATION:
;    APPLICANT:  SMITH, DOUGLAS
;    APPLICANT:  MAO, JEN-I
;    TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES
;    TITLE OF INVENTION:  RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
;    TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;    NUMBER OF SEQUENCES:  411
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  WOLF, GREENFIELD & SACKS, P.C.
;      STREET:  600 ATLANTIC AVENUE
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/311,731A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GATES, EDWARD R.
;      REGISTRATION NUMBER:  31,616
;      REFERENCE/DOCKET NUMBER:  C0044/7125
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/720-3500
;      TELEFAX:  617/720-2441
;  INFORMATION FOR SEQ ID NO:  158:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  618 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium leprae
US-08-311-731A-158

  Query Match             5.5%;  Score 83.5;  DB 4;  Length 618;
  Best Local Similarity   24.1%;  Pred. No. 4;
  Matches   60;  Conservative   30;  Mismatches  102;  Indels   57;  Gaps   14;

Qy          4 RGTGCSAGLLMTVGWL---LLAGLQSARG-TNVTAAVQDAGLAHEGEGEE-ETENNDSET 58
              |      ||:  | |      |:| :: || | |:         | |       ||   ::
Db         58 RALANQAGVKGTSGMRKSELIAAIEECRGQANGTSVNDGPSRDHGGSATAISTEALAAQE 117

Qy         59 AENYAPPETEDVSNRNVVKEVEFGMCTVTCGIGVREVIL----------TNGCPGGESKC 108
              :|||  |      |  :| :    || |    |               |   ||:|
Db        118 EQNYAIVEVSRRERRGASREAD-----VTAGTSTAEATESDCQGTADDDTRTLQGGQSD- 171

Qy        109 VVRVEECRGPTDCGWGKPISESLESVRLACIHTSPL----------NRPKYMWKLLR 155
                : || ||| | |  : : :   |::       |            ||:  :  |
Db        172 -TKTEE-RGP-DVGNDQGVEQQSSSLQ-------PRGDDDGEGRQGRRGRRFRDRDRRRR 221

Qy        156 QDQQSIILVNDSAILEVRKES--HPLAFECDTLDNNEIVATIKF-----TVYTSSELQMR 208
              ::       | |  |:|::    |:|   | |||   | |  :      || |  : :|
Db        222 GERS-----GDGAEAELRQDDVVQPVAGILDVLDNYAFVRTSGYLAGPHDVYVSMSM-VR 275

Qy        209 RSSLPATDA 217
              :: |   ||
Db        276 KNGLRRGDA 284


RESULT 8
US-09-248-796A-14534
; Sequence 14534, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14534
;   LENGTH: 519
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-14534

  Query Match             5.4%;  Score 83;  DB 4;  Length 519;
  Best Local Similarity   22.9%;  Pred. No. 3.5;
  Matches   30;  Conservative   30;  Mismatches   47;  Indels   24;  Gaps    5;

Qy        127 ISESLESVRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTL 186
              : |:||:  :  :  | :|:  :     :|:::| :  |::                  |
Db         25 LRETLETAIIISVLLSFINQRSH-----KQEEESTLSTNNNN-------------NNDNN 66

Qy        187 DNNEIVATIKFTVYTSSELQMRRSSLPATDAALIFV-LTIGVIIC--VFIIFLLIFIIIN 243
              |||||    :  |:  |  | |:: :       :::   :|:|||  : |||:|||  :
Db         67 DNNEIENENENAVHPISISQSRKADMNRKLKFQVWIGAILGLIICFIIGIIFVLIFYFVG 126

Qy        244 ---WAAVKAFW 251
                 |:  :  |
Db        127 QDYWSYTERVW 137


RESULT 9
US-09-134-001C-4799
; Sequence 4799, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4799
;   LENGTH: 366
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4799

  Query Match             5.4%;  Score 82.5;  DB 3;  Length 366;
  Best Local Similarity   19.0%;  Pred. No. 2.3;
  Matches   47;  Conservative   41;  Mismatches   80;  Indels   79;  Gaps   11;

Qy         21 LAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDS--------ETAENYAPPETEDVSN 72
              : |:      ::|  ::: |:   |      |:| |:         :|||      | : ||
Db        107 IPGISGVDTRSITRKIRNYGVLRAG----FTDNKDNIQELVEQLKTAE-LPRDEVQTVST 161

Qy         73 R--------------------NVVKEVEFGMCTVT-----------CGIGVREVILTNG 100
              :                    |:|:|:    | ||             |:    |:|:||
Db        162 KTPYVSTGSDLSVVLLDFGKKQNIVRELNLRGCNVTVVPYDTSAEEILGMSPDGVMLSNG 221

Qy        101 CPGGESKCVVRVEECRGPTDCGWGK------------PISESLESVRLACIHTSPLNRF 147
               ||   :  | ::  ||      ||                :|:    | ::    |
Db        222 -PGDPDEVDVALDMIRGIL----GKIPFFGICLGHQLFALSQGATSFKMKFGHRGA---- 272

Qy        148 KYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVATIKFTVYTSSELQM 207
              :  | ||  :           :::  ::|  : :||:| | ::  |    :   :   :
Db        273 NHPVKDLRTGK----------IDITSQNHGYSIDCDSLKNTDLEVT-HIALNDGTVEGL 320

Qy        208 RRSSLPA 214
              |   |||
Db        321 RHKELPA 327


RESULT 10
```

```
US-09-513-783A-142
; Sequence 142, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
;  APPLICANT:  Giuliano, Kenneth A.
;  APPLICANT:  Kapur, Ravi
;  TITLE OF INVENTION:  A System for Cell Based Screening
;  FILE REFERENCE:  97-022-L1
;  CURRENT APPLICATION NUMBER: US/09/513,783A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS:  180
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 142
;    LENGTH: 566
;    TYPE: PRT
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Description of Artificial Sequence: Size exclusion
;    OTHER INFORMATION: target sequence
US-09-513-783A-142

  Query Match             5.2%;  Score 80;  DB 4;  Length 566;
  Best Local Similarity   22.1%;  Pred. No. 8.7;
  Matches   53;  Conservative   32;  Mismatches  101;  Indels   54;  Gaps    7;

Qy          2 SPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAEN 61
              ||: |  :  : |      ||  :    |  |      |:    | ||    ||  :||
Db        344 SPKETETTLPIKMD-----LAPPEDVLLTKETELAPAKGMVSLSEIEEALAKNDVRSAEI 398

Qy         62 YAPPETEDVSNRNVVKEVEPGMCT--------------VTCGIGVREVILTNGCPGGES 106
                ||       || | |  : |                ||  :     ::|:  |  |:
Db        399 PVAQET-------VVSETEVVLATEVVLPSDPITTLTKDVTLPLEAERPLVTDMTPSLET 451

Qy        107 KCVVRVEECRGPTDCGWGK-------PISESLESVRLACIHTSPLNRFKYMWKLLRQDQQ 159
              :  :  :|   ||:   |         | ||      :  :  : :  |  :  |  ::
Db        452 EMTLG-KETAPPTETNLGMAKDMSPLPESEVTLGKDVVILPETKVAEFNNVTPLSEEEVT 510

Qy        160 SIILVNDSAILEVRKESHPLAFECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAAL 219
              |:  :: ||  |      |||   |             ::: :|| :  |  ||:| ||
Db        511 SVKDMSPSAETEA-----PLAKNAD-------------LHSGTELIVDNSMAPASDLAL 551


RESULT 11
US-09-430-656-142
; Sequence 142, Application US/09430656
; Patent No. 6756207
; GENERAL INFORMATION:
;  APPLICANT: Giuliano, Kenneth A.
;  APPLICANT:  Bright, Gary
;  APPLICANT:  Olson, Keith
;  APPLICANT:  Burroughs-Tencza, Sarah
;  TITLE OF INVENTION: A System for Cell Based Screening
;  FILE REFERENCE: 97-022-K
;  CURRENT APPLICATION NUMBER: US/09/430,656
;  CURRENT FILING DATE:  1999-10-29
;  PRIOR APPLICATION NUMBER: 09/398,965
;  PRIOR FILING DATE: 1999-09-17
;  PRIOR APPLICATION NUMBER: 09/031,271
;  PRIOR FILING DATE: 1998-02-27
;  PRIOR APPLICATION NUMBER: 08/810,983
;  PRIOR FILING DATE: 1997-02-27
;  PRIOR APPLICATION NUMBER: 60/136,078
;  PRIOR FILING DATE: 1999-05-26
;  PRIOR APPLICATION NUMBER: 60/106,308
;  PRIOR FILING DATE: 1998-10-30
;  NUMBER OF SEQ ID NOS: 168
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 142
;    LENGTH: 566
;    TYPE: PRT
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Description of Artificial Sequence: Size exclusion
;    OTHER INFORMATION: target sequence
US-09-430-656-142

  Query Match             5.2%;  Score 80;  DB 4;  Length 566;
  Best Local Similarity   22.1%;  Pred. No. 8.7;
  Matches   53;  Conservative   32;  Mismatches  101;  Indels   54;  Gaps    7;

Qy          2 SPRGTGCSAGLLMTVGWLLLAGLQSARGTNVTAAVQDAGLAHEGEGEEETENNDSETAEN 61
              ||: |  :  : |      ||  :    |  |      |:    | ||    ||  :||
Db        344 SPKETETTLPIKMD-----LAPPEDVLLTKETELAPAKGMVSLSEIEEALAKNDVRSAEI 398

Qy         62 YAPPETEDVSNRNVVKEVEPGMCT--------------VTCGIGVREVILTNGCPGGES 106
                ||       || | |  : |                ||  :     ::|:  |  |:
Db        399 PVAQET-------VVSETEVVLATEVVLPSDPITTLTKDVTLPLEAERPLVTDMTPSLET 451

Qy        107 KCVVRVEECRGPTDCGWGK-------PISESLESVRLACIHTSPLNRFKYMWKLLRQDQQ 159
              :  :  :|   ||:   |         | ||      :  :  : :  |  :  |  ::
Db        452 EMTLG-KETAPPTETNLGMAKDMSPLPESEVTLGKDVVILPETKVAEFNNVTPLSEEEVT 510

Qy        160 SIILVNDSAILEVRKESHPLAFECDTLDNNEIVATIKFTVYTSSELQMRRSSLPATDAAL 219
              |:  :: ||  |      |||   |             ::: :|| :  |  ||:| ||
Db        511 SVKDMSPSAETEA-----PLAKNAD-------------LHSGTELIVDNSMAPASDLAL 551


RESULT 12
US-07-885-972A-2
; Sequence 2, Application US/07885972A
; Patent No. 5460950
;  GENERAL INFORMATION:
;    APPLICANT:  Barr, Philip J.
;    APPLICANT:  Brake, Anthony J.
;    APPLICANT:  Kaufman, Rnadal J.
;    APPLICANT:  Tekamp-Olson, Patricia
;    APPLICANT:  Wasley, Louise
;    APPLICANT:  Wong, Polly A.
;    TITLE OF INVENTION:  Expression of PACE in Host Cells and
;    TITLE OF INVENTION:  Methods of Use Thereof
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson & Howson
;      STREET:  Spring House Corporate Center, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,972A
;      FILING DATE:  19920520
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,092
;      FILING DATE:  26-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/620,859
;      FILING DATE:  29-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,443
;      FILING DATE:  29-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,457
;      FILING DATE:  30-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GI5181A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
```

```
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  794 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-885-972A-2

  Query Match             5.2%;  Score 80;  DB 1;  Length 794;
  Best Local Similarity   21.7%;  Pred. No. 15;
  Matches   75;  Conservative   27;  Mismatches  111;  Indels  132;  Gaps   18;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGL---QSARGTNVTAAVQDAGLAHEGEGEEETENNDSE 57
              :|| ||  :        |||      || | |  | :       :  ||   |  ::
Db        506 VSPMGTRST----------LLAARPHDYSADGFNDWAFMTTHSWDEDPSGEWVLEIENTS 555

Qy         58 TAENY------------------APPE-----TEDVSNRNVVKEVEFGM----CTVTCG 89
               | ||                 |||      |   |   || |  | :    |   |
Db        556 EANNYGTLTKFTLVLYGTAPEGLPVPPESSGCKTLTSSQACVVCEEGFSLHQKSCVQHCP 615

Qy         90 IGVR--------------EVILTNGCPGGESKCVVRVEECRGP--TDCGWGKPISESLES 133
               |                | |  : |    : |       |:||  |||     |   ||:
Db        616 PGFAPQVLDTHYSTENDVETIRASVCAPCHASCAT----CQGPALTDC-LSCPSHASLDP 670

Qy        134 VRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA 193
              |  |          || | |                :|| |             :
Db        671 VEQTC--------------SRQSQSS------------RESPP----------QQQPP 692

Qy        194 TIKFTVYTSSELQ--MRRSSLPATDAALIFVLTIGVIICVFI--IFLLIFIIINWAAVKA 249
              :   |      |:  :  | ||   | |          | ||  :|: :|:::   :  :
Db        693 RLPPEVEAGQRLRAGLLPSHLPEVVAGL---------SCAFIVLVFVTVFLVLQLRSGFS 743

Qy        250 FWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              | | |  |     ::  : ||       || |          | |
Db        744 FRGVKVYT----MDRGLISYKG------LPPE---------AWQE 769


RESULT 13
US-07-885-972A-4
; Sequence 4, Application US/07885972A
; Patent No. 5460950
;  GENERAL INFORMATION:
;    APPLICANT:  Barr, Philip J.
;    APPLICANT:  Brake, Anthony J.
;    APPLICANT:  Kaufman, Rnadal J.
;    APPLICANT:  Tekamp-Olson, Patricia
;    APPLICANT:  Wasley, Louise
;    APPLICANT:  Wong, Polly A.
;    TITLE OF INVENTION:  Expression of PACE in Host Cells and
;    TITLE OF INVENTION:  Methods of Use Thereof
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson & Howson
;      STREET:  Spring House Corporate Center, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,972A
;      FILING DATE:  19920520
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,092
;      FILING DATE:  26-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/620,859
;      FILING DATE:  29-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,443
;      FILING DATE:  29-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,457
;      FILING DATE:  30-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GI5181A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  794 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-885-972A-4

  Query Match             5.2%;  Score 80;  DB 1;  Length 794;
  Best Local Similarity   21.7%;  Pred. No. 15;
  Matches   75;  Conservative   27;  Mismatches  111;  Indels  132;  Gaps   18;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGL---QSARGTNVTAAVQDAGLAHEGEGEEETENNDSE 57
              :|| ||  :        |||      || | |  | :       :  ||   |  ::
Db        506 VSPMGTRST----------LLAARPHDYSADGFNDWAFMTTHSWDEDPSGEWVLEIENTS 555

Qy         58 TAENY------------------APPE-----TEDVSNRNVVKEVEFGM----CTVTCG 89
               | ||                 |||      |   |   || |  | :    |   |
Db        556 EANNYGTLTKFTLVLYGTAPEGLPVPPESSGCKTLTSSQACVVCEEGFSLHQKSCVQHCP 615

Qy         90 IGVR--------------EVILTNGCPGGESKCVVRVEECRGP--TDCGWGKPISESLES 133
               |                | |  : |    : |       |:||  |||     |   ||:
Db        616 PGFAPQVLDTHYSTENDVETIRASVCAPCHASCAT----CQGPALTDC-LSCPSHASLDP 670

Qy        134 VRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA 193
              |  |          || | |                :|| |             :
Db        671 VEQTC--------------SRQSQSS------------RESPP----------QQQPP 692

Qy        194 TIKFTVYTSSELQ--MRRSSLPATDAALIFVLTIGVIICVFI--IFLLIFIIINWAAVKA 249
              :   |      |:  :  | ||   | |          | ||  :|: :|:::   :  :
Db        693 RLPPEVEAGQRLRAGLLPSHLPEVVAGL---------SCAFIVLVFVTVFLVLQLRSGFS 743

Qy        250 FWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              | | |  |     ::  : ||       || |          | |
Db        744 FRGVKVYT----MDRGLISYKG------LPPE---------AWQE 769


RESULT 14
US-08-865-203-2
; Sequence 2, Application US/08865203
; Patent No. 5935815
;  GENERAL INFORMATION:
;    APPLICANT:  van de Ven, Willem Jan Marie
;    APPLICANT:  van den Ouweland, Anna Maria Wilhelmina
;    APPLICANT:  Van Duijnhoven, Johannes Lambertus Petrus
;    APPLICANT:  Robroek, Antonius Johannes Maria
;    APPLICANT:  Koning, Piet Nico Maria
;    TITLE OF INVENTION:  Pharmaceutical Composition Having An
;    TITLE OF INVENTION:  Endoproteolytic Activity; A Process for
;    TITLE OF INVENTION:  Endoproteolytically Processing (Precursor)
;    TITLE OF INVENTION:  Proteins And For The (Micro)Biological
;    TITLE OF INVENTION:  Production Of Proteins
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HOFFMANN & BARON, LLP
;      STREET:  350 Jericho Turnpike
;      CITY:  Jericho
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  11753
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/865,203
;      FILING DATE:  29-MAY-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tran, Jessica H.
;      REGISTRATION NUMBER:  40,846
;      REFERENCE/DOCKET NUMBER:  294-41 DIV II
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (516) 822-3550
;      TELEFAX:  (516) 822-3582
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  794 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-865-203-2

  Query Match             5.2%;  Score 80;  DB 2;  Length 794;
  Best Local Similarity   21.7%;  Pred. No. 15;
  Matches   75;  Conservative   27;  Mismatches  111;  Indels  132;  Gaps   18;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGL---QSARGTNVTAAVQDAGLAHEGEGEEETENNDSE 57
              :|| ||  :           |||       || | |  | :       :  ||   |  ::
Db        506 VSPMGTRST----------LLAARPHDYSADGFNDWAFMTTHSWDEDPSGEWVLEIENTS 555

Qy         58 TAENY------------------APPE-----TEDVSNRNVVKEVEFGM----CTVTCG 89
               | ||                  |||      |   |  || | | :    |   |
Db        556 EANNYGTLTKFTLVLYGTAPEGLPVPPESSGCKTLTSSQACVVCEEGFSLHQKSCVQHCP 615

Qy         90 IGVR-------------EVILTNGCPGGESKCVVRVEECRGP--TDCGWGKPISESLES 133
               |                | |  : |    : |      |:||  |||     |   ||:
Db        616 PGFAPQVLDTHYSTENDVETIRASVCAPCHASCAT----CQGPALTDC-LSCPSHASLDP 670

Qy        134 VRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA 193
              |  |                 || | |            :|| |           :
Db        671 VEQTC--------------SRQSQSS------------RESPP----------QQQPP 692

Qy        194 TIKFTVYTSSELQ--MRRSSLPATDAALIFVLTIGVIICVFI--IFLLIFIIINWAAVKA 249
              :   |    |:  :  | ||   | |           |.||  :|: :|:::   :  :
Db        693 RLPPEVEAGQRLRAGLLPSHLPEVVAGL---------SCAFIVLVFVTVFLVLQLRSGFS 743

Qy        250 FWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              | | |  |     ::  : ||       || |          | |
Db        744 FRGVKVYT----MDRGLISYKG------LPPE---------AWQE 769


RESULT 15
US-08-745-880-2
; Sequence 2, Application US/08745880
; Patent No. 5965425
;  GENERAL INFORMATION:
;    APPLICANT:  Barr, Philip J.
;    APPLICANT:  Brake, Anthony J.
;    APPLICANT:  Kaufman, Rnadal J.
;    APPLICANT:  Tekamp-Olson, Patricia
;    APPLICANT:  Wasley, Louise
;    APPLICANT:  Wong, Polly A.
;    TITLE OF INVENTION:  Expression of PACE in Host Cells and
;    TITLE OF INVENTION:  Methods of Use Thereof
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson & Howson
;      STREET:  Spring House Corporate Center, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/745,880
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/480,382
;      FILING DATE:  07-JUN-1995
;      APPLICATION NUMBER:  US 07/885,972
;      FILING DATE:  20-MAY-1992
;      APPLICATION NUMBER:  US 07/621,092
;      FILING DATE:  26-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/620,859
;      FILING DATE:  29-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,443
;      FILING DATE:  29-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,457
;      FILING DATE:  30-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GI5181A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  794 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-745-880-2

  Query Match             5.2%;  Score 80;  DB 2;  Length 794;
  Best Local Similarity   21.7%;  Pred. No. 15;
  Matches   75;  Conservative   27;  Mismatches  111;  Indels  132;  Gaps   18;

Qy          1 MSPRGTGCSAGLLMTVGWLLLAGL---QSARGTNVTAAVQDAGLAHEGEGEEETENNDSE 57
              :|| ||  :           |||       || | |  | :       :  ||   |  ::
Db        506 VSPMGTRST----------LLAARPHDYSADGFNDWAFMTTHSWDEDPSGEWVLEIENTS 555

Qy         58 TAENY------------------APPE-----TEDVSNRNVVKEVEFGM----CTVTCG 89
               | ||                  |||      |   |  || | | :    |   |
Db        556 EANNYGTLTKFTLVLYGTAPEGLPVPPESSGCKTLTSSQACVVCEEGFSLHQKSCVQHCP 615

Qy         90 IGVR-------------EVILTNGCPGGESKCVVRVEECRGP--TDCGWGKPISESLES 133
               |                | |  : |    : |      |:||  |||     |   ||:
Db        616 PGFAPQVLDTHYSTENDVETIRASVCAPCHASCAT----CQGPALTDC-LSCPSHASLDP 670

Qy        134 VRLACIHTSPLNRFKYMWKLLRQDQQSIILVNDSAILEVRKESHPLAFECDTLDNNEIVA 193
              |  |                 || | |            :|| |           :
Db        671 VEQTC--------------SRQSQSS------------RESPP----------QQQPP 692

Qy        194 TIKFTVYTSSELQ--MRRSSLPATDAALIFVLTIGVIICVFI--IFLLIFIIINWAAVKA 249
              :   |    |:  :  | ||   | |           |.||  :|: :|:::   :  :
Db        693 RLPPEVEAGQRLRAGLLPSHLPEVVAGL---------SCAFIVLVFVTVFLVLQLRSGFS 743

Qy        250 FWGAKASTPEVQSEQSSVRYKDSTSLDQLPTEMPGEDDALSEWNE 294
              | | |  |     ::  : ||       || |          | |
Db        744 FRGVKVYT----MDRGLISYKG------LPPE---------AWQE 769


Search completed: April 15, 2005, 14:16:29
```

Job time : 24 secs